GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:46 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcatttcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.2	76.8	265	163	BE125672	BE125672 DG1_54_E0
	2	19.2	76.8	353	114	AW287736	AW287736 LG1_271_G
	3	19.2	76.8	542	166	BE357782	BE357782 DG1_22_F0
c	4	18.2	72.8	372	189	W00149	W00149 TgESTzy77e0
c	5	18.2	72.8	431	122	AW925187	AW925187 WS1_76_E0
c	6	18.2	72.8	544	239	AZ143198	AZ143198 SP_0028_B
	7	18.2	72.8	613	228	AQ438224	AQ438224 HS_5083_B
c	8	17.8	71.2	538	30	AV439422	AV439422 AV439422
c	9	17.6	70.4	508	227	AQ358636	AQ358636 HS_5030_A
	10	17.6	70.4	574	167	BE405752	BE405752 WHE1214_H
c	11	17.6	70.4	576	163	BE131409	BE131409 L48-1384T
	12	17.6	70.4	1032	168	BF682210	BF682210 602117637
c	13	17.2	68.8	325	137	BE595789	BE595789 PI1_54_A0
c	14	17.2	68.8	343	118	AW604656	AW604656 CM0-CT031
	15	17.2	68.8	349	150	BF524069	BF524069 UI-R-C0-h
	16	17.2	68.8	364	218	AF116319	AF116319 AF116319
c	17	17.2	68.8	654	14	AA946503	AA946503 EST202002
	18	17.2	68.8	719	122	AW940604	AW940604 GH15883.3
	19	17.2	68.8	988	222	CNS05JPQ	AL340487 Tetraodon
	20	17.2	68.8	997	222	CNS05BPT	AL330122 Tetraodon
	21	17.2	68.8	1081	222	CNS05R7O	AL350205 Tetraodon
	22	17.2	68.8	1095	221	CNS04VWF	AL309624 Tetraodon
	23	17	68.0	113	7	AA422533	AA422533 vf14a04.s
	24	17	68.0	300	148	BF386481	BF386481 UI-R-CA1-
c	25	17	68.0	332	24	AI724043	AI724043 RHIZ1_25_
	26	17	68.0	360	250	AZ817869	AZ817869 2M0087N04
c	27	17	68.0	421	120	AW747232	AW747232 WS1_66_E0
	28	17	68.0	432	148	BF387098	BF387098 UI-R-CA1-
c	29	17	68.0	446	175	BG241150	BG241150 OV1_38_H0
c	30	17	68.0	456	18	AI283774	AI283774 qj75e06.x
c	31	17	68.0	463	175	BG240831	BG240831 OV1_38_H0
c	32	17	68.0	471	173	BG047654	BG047654 OV1_26_E0
c	33	17	68.0	473	173	BG048007	BG048007 OV1_26_E0
	34	17	68.0	476	148	BF388049	BF388049 UI-R-CA1-
c	35	17	68.0	489	122	AW923923	AW923923 WS1_30_A0
c	36	17	68.0	497	166	BE355471	BE355471 DG1_114_H
c	37	17	68.0	502	120	AW747172	AW747172 WS1_66_E0
c	38	17	68.0	506	244	AZ475043	AZ475043 1M0292H17
	39	17	68.0	510	148	BF390169	BF390169 UI-R-CA1-
c	40	17	68.0	511	166	BE364076	BE364076 PI1_11_A0
c	41	17	68.0	511	166	BE364141	BE364141 PI1_11_A0
	42	17	68.0	515	235	AQ940100	AQ940100 Sheared D
c	43	17	68.0	518	166	BE355404	BE355404 DG1_114_H
c	44	17	68.0	519	166	BE355326	BE355326 DG1_114_H
	45	17	68.0	529	258	TA367G09Q	AL495254 T. brucei

ALIGNMENTS

```
RESULT   1
BE125672
LOCUS       BE125672      265 bp    mRNA            EST       19-JUL-2000
DEFINITION  DG1_54_E04.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE125672
VERSION     BE125672.1  GI:8548262
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 265
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       54 a     98 c     75 g     38 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 163;  Length 265;
  Best Local Similarity   87.5%;  Pred. No. 66;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||| || |||||||||| |||||
Db      61 CATTGCGCTCCTCGTCAGCTGCAG 84


RESULT   2
AW287736
LOCUS       AW287736      353 bp    mRNA            EST       19-JUL-2000
DEFINITION  LG1_271_G07.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW287736
VERSION     AW287736.2  GI:6859727
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6677580.
            Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 352
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
```

seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI ; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 55 a 120 c 123 g 55 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 114; Length 353;
Best Local Similarity 87.5%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
|||| || |||||||||| |||||
Db 47 CATTGCGCTCCTCGTCAGCTGCAG 70

RESULT 3
BE357782
LOCUS BE357782 542 bp mRNA EST 20-JUL-2000
DEFINITION DG1_22_F01.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE357782
VERSION BE357782.1 GI:9299339
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 542)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 537
POLYA=No.
FEATURES Location/Qualifiers
source 1. .542
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 80 a 196 c 187 g 79 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 166; Length 542;
Best Local Similarity 87.5%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
|||| || |||||||||| |||||
Db 61 CATTGCGCTCCTCGTCAGCTGCAG 84

RESULT 4
W00149/c
LOCUS W00149 372 bp mRNA EST 22-MAY-2000
DEFINITION TgESTzy77e07.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone tgzy77e07.r1 5', mRNA sequence.
ACCESSION W00149
VERSION W00149.1 GI:1271572
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 372)
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A. , Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L., Waterston,R.H. and Boothroyd,J.
TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
David Sibley at toxoest@borcim.wustl.edu for further information relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 321.
FEATURES Location/Qualifiers
source 1. .372
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgzy77e07.r1"
/clone_lib="TgRH Tachyzoite cDNA"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was constructed by K.L. Wan, Cambridge University. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the Lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."
BASE COUNT 103 a 111 c 85 g 72 t 1 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 189; Length 372;
Best Local Similarity 87.0%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atttcgatcctcgtcaggtgcag 25
|||||| || ||||| |||||||
Db 208 ATTTCGTTCATCGTCTGGTGCAG 186

RESULT 5
AW925187/c
LOCUS AW925187 431 bp mRNA EST 19-JUL-2000
DEFINITION WS1_76_E02.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION AW925187
VERSION AW925187.1 GI:8091013
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 431)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: water-stressed plants

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 339
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       98 a    135 c    106 g     92 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 122;  Length 431;
  Best Local Similarity   87.0%;  Pred. No. 2.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgca 24
          ||||  |||||||||||||||| |
Db    425 CATTGGGATCCTCGTCAGGTGGA 403


RESULT   6
AZ143198/c
LOCUS       AZ143198      544 bp    DNA             GSS       28-AUG-2000
DEFINITION  SP_0028_B2_H12_T7 Strongylocentrotus purpuratus, purple sea urchin,
            sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=28 Col=24 Row=P, DNA sequence.
ACCESSION   AZ143198
VERSION     AZ143198.1  GI:8295101
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 28  row: P  column: 24
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=28 Col=24 Row=P"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      145 a    116 c     87 g    190 t      6 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 239;  Length 544;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          || ||  ||||||||||||  ||||
Db    331 GCCTTGCGATCCTCGTCANATGCA 308


RESULT   7
AQ438224
LOCUS       AQ438224      613 bp    DNA             GSS       31-MAR-1999
DEFINITION  HS_5083_B2_H12_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=659 Col=24 Row=P, DNA sequence.
ACCESSION   AQ438224
VERSION     AQ438224.1  GI:4549563
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 659  row: P  column: 24
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=659 Col=24 Row=P"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      170 a     97 c    136 g    194 t     16 others
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 228;  Length 613;
  Best Local Similarity   87.0%;  Pred. No. 2.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgcag 25
           ||||| |||||| ||| ||||||
Db     158 ATTTCTATCCTCTTCATGTGCAG 180


RESULT   8
AV439422/c
LOCUS       AV439422      538 bp    mRNA            EST       23-AUG-2000
DEFINITION  AV439422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PS050c10_r 5', mRNA sequence.
ACCESSION   AV439422
VERSION     AV439422.1  GI:8594647
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS050c10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      114 a    158 c    170 g     96 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 30;  Length 538;
  Best Local Similarity   90.5%;  Pred. No. 3.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggt 21
           ||||||| |||| ||||||||
Db     371 GCATTTCCATCCCCGTCAGGT 351


RESULT   9
AQ358636/c
LOCUS       AQ358636      508 bp    DNA             GSS       06-MAR-1999
DEFINITION  HS_5030_A1_C01_T7 RPCI11 Human Male BAC Library Homo sapiens
            genomic clone Plate=606 Col=1 Row=E, DNA sequence.
ACCESSION   AQ358636
VERSION     AQ358636.1  GI:4207512
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 606  row: E  column: 1
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=606 Col=1 Row=E"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      147 a     97 c     95 g    153 t     16 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 227;  Length 508;
  Best Local Similarity   80.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||   | |||| ||||||||||||
Db      98 GCANAACTATCCGCGTCAGGTGCAG 74


RESULT  10
BE405752
LOCUS       BE405752      574 bp    mRNA            EST       21-JUL-2000
DEFINITION  WHE1214_H08_O16ZS Wheat etiolated seedling root cDNA library
            Triticum aestivum cDNA clone WHE1214_H08_O16, mRNA sequence.
ACCESSION   BE405752
VERSION     BE405752.1  GI:9365220
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Strategene SK primer.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1214_H08_O16"
```

/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

BASE COUNT 155 a 125 c 153 g 141 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 167; Length 574;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgca 24
|||||||| | | ||||||| |||
Db 289 GCATTTCGCTGCCCGTCAGGGGCA 312

RESULT 11
BE131409/c
LOCUS BE131409 576 bp mRNA EST 20-FEB-2001
DEFINITION L48-1384T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1384, mRNA sequence.
ACCESSION BE131409
VERSION BE131409.1 GI:8578772
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 576)
AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR PRimers
FORWARD: T7
BACKWARD: T3
Plate: L48-14 row: G column: 12
Seq primer: T3
High quality sequence stop: 350
POLYA=No.
FEATURES Location/Qualifiers
source 1. .576
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L48-1384"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library
, 48 hours NaCl treatment"
/tissue_type="Leaf, 48 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 152 a 141 c 118 g 165 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 163; Length 576;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
|||||||| | ||||| | |||||
Db 465 CATTTCGACCATCGTCTGCTGCAG 442

RESULT 12
BF682210
LOCUS BF682210 1032 bp mRNA EST 21-DEC-2000
DEFINITION 602117637T1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone IMAGE:3468776 3', mRNA sequence.
ACCESSION BF682210
VERSION BF682210.1 GI:11956105
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1032)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: M. Bento Soares, Ph.D.
cDNA Library Preparation: Bento Soares and M. Fatima Bonaldo
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8476 row: j column: 09
High quality sequence start: 21
High quality sequence stop: 789.
FEATURES Location/Qualifiers
source 1. .1032
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3468776"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 288 a 261 c 218 g 265 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 168; Length 1032;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
|||||| || ||| | ||||||||
Db 810 CATTTCTATTCTCCTTAGGTGCAG 833

RESULT 13
BE595789/c

```
LOCUS       BE595789      325 bp    mRNA            EST       18-AUG-2000
DEFINITION  PI1_54_A07.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE595789
VERSION     BE595789.1  GI:9850862
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 254
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT      104 a     62 c     78 g     81 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 137;  Length 325;
  Best Local Similarity   86.4%;  Pred. No. 6e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtg 22
           |||||| ||||||| ||||| |
Db     148 GCATTTTGATCCTCATCAGGAG 127


RESULT  14
AW604656/c
LOCUS       AW604656      343 bp    mRNA            EST       23-MAR-2000
DEFINITION  CM0-CT0312-190100-153-d11 CT0312 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW604656
VERSION     AW604656.1  GI:7309397
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM0&t2=CM0-CT0312-
            190100-153-d11&t3=2000-01-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 116.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0312"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       77 a    101 c     92 g     73 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 118;  Length 343;
  Best Local Similarity   86.4%;  Pred. No. 6.1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           ||||||| || || ||||||||
Db     252 CATTTCGGTCATCTTCAGGTGC 231


RESULT  15
BF524069
LOCUS       BF524069      349 bp    mRNA            EST       11-DEC-2000
DEFINITION  UI-R-C0-hs-f-07-0-UI.r1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-hs-f-07-0-UI 5', mRNA sequence.
ACCESSION   BF524069
VERSION     BF524069.1  GI:11632036
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
```

```
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1792330
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .349
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-C0-hs-f-07-0-UI"
                      /clone_lib="UI-R-C0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                      library is a subtracted library derived from the UI-R-A1
                      and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                      mixture of individually tagged normalized libraries
                      constructed from rat  placenta, adult lung, brain, liver,
                      kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                      library  consisted of a mixture of individually tagged
                      normalized libraries constructed from 8, 12 and 18-day
                      embryo. The tag is a string of 3-5 nucleotides present
                      between the Not I  site and the oligo-dT track which
                      allows identification of the library of origin of a clone
                      within the mixture. The subtracted library (UI-R-C0) was
                      constructed as follows: PCR amplified cDNA inserts from a
                      pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                      been derived was used as a driver in a hybridization with
                      the pooled UI-R-A1 and UI-R-E1 library in the form of
                      single-stranded circles.  The remaining single-stranded
                      circles (subtracted library) was purified by
                      hydroxyapatite column chromatography, converted to
                      double-stranded circles and electroporated into DH10B
                      bacteria (Life Technologies) to generate the UI-R-C0
                      library. This procedure has been previously described
                      (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                      1996)"
BASE COUNT       76 a    105 c     95 g     73 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 150;  Length 349;
  Best Local Similarity   86.4%;  Pred. No. 6.1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgc 23
          || ||| |||||||||||| ||
Db    295 CACTTCCATCCTCGTCAGGGGC 316


RESULT  16
AF116319
LOCUS       AF116319      364 bp    DNA             GSS       12-APR-2001
DEFINITION  AF116319 Armillaria gallica W40f Armillaria gallica genomic, DNA
            sequence.
ACCESSION   AF116319
VERSION     AF116319.1  GI:4761841
KEYWORDS    GSS.
SOURCE      Armillaria gallica.
  ORGANISM  Armillaria gallica
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Hymenomycetidae;
            Agaricales; Tricholomataceae; Armillaria.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Hodnett,B. and Anderson,J.B.
  TITLE     Genomic stability of two individuals of Armillaria gallica
  JOURNAL   Mycologia 92 (5), 894-899 (2000)
COMMENT     Contact: Anderson JB
            Biology
            University of Toronto at Mississauga
            3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada
            Email: bhodnett@credit.erin.utoronto.ca
            Class: unknown.
FEATURES              Location/Qualifiers
     source           1. .364
                      /organism="Armillaria gallica"
                      /strain="W40f"
                      /db_xref="taxon:47427"
                      /clone_lib="Armillaria gallica W40f"
BASE COUNT       56 a    107 c     71 g    128 t      2 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 218;  Length 364;
  Best Local Similarity   86.4%;  Pred. No. 6.1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgc 23
          ||||||| |||  |||||||||
Db     80 CATTTCGTTCCAAGTCAGGTGC 101


RESULT  17
AA946503/c
LOCUS       AA946503      654 bp    mRNA            EST       08-JAN-1999
DEFINITION  EST202002 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVAR89 3' end, mRNA sequence.
ACCESSION   AA946503
VERSION     AA946503.1  GI:4132826
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On May 1, 1998 this sequence version replaced gi:3106419.
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .654
                      /organism="Rattus sp."
                      /db_xref="taxon:10118"
                      /clone="ROVAR89"
                      /clone_lib="Normalized rat ovary, Bento Soares"
                      /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                      Site_2: NotI"
BASE COUNT      151 a    143 c    198 g    162 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 14;  Length 654;
  Best Local Similarity   86.4%;  Pred. No. 6.4e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgc 23
          || ||| |||||||||||| ||
Db    575 CACTTCCATCCTCGTCAGGGGC 554


RESULT  18
AW940604
LOCUS       AW940604      719 bp    mRNA            EST       30-MAY-2000
DEFINITION  GH15883.3prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15883 3, mRNA sequence.
```

```
ACCESSION   AW940604
VERSION     AW940604.1  GI:8116050
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: GH15883.5prime
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed.
            Plate: GH.158  row: G  column: 11
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15883"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      162 a    196 c    164 g    197 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 122;  Length 719;
  Best Local Similarity   86.4%;  Pred. No. 6.5e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgca 24
           ||| |||| |||| ||||||||
Db     441 ATTGCGATACTCGCCAGGTGCA 462


RESULT  19
CNS05JPQ
LOCUS       CNS05JPQ      988 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            028C23 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL340487
VERSION     AL340487.1  GI:8234245
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 988)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 988)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .988
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="028C23"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA028AB12A2~end : T3"
BASE COUNT      281 a    168 c    175 g    331 t     33 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 222;  Length 988;
  Best Local Similarity   86.4%;  Pred. No. 6.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgcag 25
           ||||| ||||| |||| |||||
Db     796 TTTCGCTCCTCTTCAGTTGCAG 817


RESULT  20
CNS05BPT
LOCUS       CNS05BPT      997 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            002M10 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL330122
VERSION     AL330122.1  GI:8223744
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 997)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .997
```

```
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="002M10"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA002BG05C1~end : T7"
BASE COUNT       260 a    222 c    193 g    296 t     26 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 222;  Length 997;
  Best Local Similarity   86.4%;  Pred. No. 6.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tttcgatcctcgtcaggtgcag 25
          ||||| ||||| |||| |||||
Db     71 TTTCGCTCCTCTTCAGTTGCAG 92


RESULT  21
CNS05R7O
LOCUS       CNS05R7O     1081 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            043G20 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL350205
VERSION     AL350205.1  GI:8243975
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1081)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1081)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1081)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1081
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="043G20"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA043BD10C1~end : T7"
BASE COUNT       350 a    186 c    188 g    342 t     15 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 222;  Length 1081;
  Best Local Similarity   86.4%;  Pred. No. 6.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tttcgatcctcgtcaggtgcag 25
          ||||| ||||| |||| |||||
Db    776 TTTCGCTCCTCTTCAGTTGCAG 797


RESULT  22
CNS04VWF
LOCUS       CNS04VWF     1095 bp    DNA             GSS       24-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            022N08 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL309624
VERSION     AL309624.1  GI:8216563
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1095)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1095)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="022N08"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA022DG04A1~end : T3"
BASE COUNT       277 a    244 c    251 g    317 t      6 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 221;  Length 1095;
  Best Local Similarity   86.4%;  Pred. No. 6.7e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tttcgatcctcgtcaggtgcag 25
          ||| |||||| |||||| ||||
Db    596 TTTAGATCCTAGTCAGGAGCAG 617


RESULT  23
AA422533
LOCUS       AA422533      113 bp    mRNA            EST       16-OCT-1997
DEFINITION  vf14a04.s1 Knowles Solter mouse unfertilized egg Mus musculus cDNA
            clone IMAGE:835662 5' similar to SW:MVIN_SALTY P37169 VIRULENCE
            FACTOR MVIN. ;, mRNA sequence.
ACCESSION   AA422533
VERSION     AA422533.1  GI:2101353
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 113)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:495878
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .113
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:835662"
                     /clone_lib="Knowles Solter mouse unfertilized egg"
                     /tissue_type="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Organ: unfertilized egg; Vector: pBluescribe
                     (modified); Site_1: MluI; Site_2: SalI; Cloned
                     unidirectionally from mRNA prepared from 5000 unfertilized
                     eggs.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs were
                     cloned into the MluI/SalI sites of a modified pBluescribe
                     vector using commercial linkers (NEB).  Average insert
                     size:  1.0 kb."
BASE COUNT       15 a     32 c     36 g     30 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 7;  Length 113;
  Best Local Similarity   80.0%;  Pred. No. 6.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||||  ||||| || ||||||||| |
Db      9 GCATGCCGATCTTCTTCAGGTGCGG 33


RESULT  24
BF386481
LOCUS       BF386481      300 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbe-e-03-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbe-e-03-0-UI 3', mRNA sequence.
ACCESSION   BF386481
VERSION     BF386481.1  GI:11371346
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bbe-e-03-0-UI"
                     /clone_lib="UI-R-CA1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex,  corpus striatum, testis, and
                     hippocampus.  For a detailed description of the  library
                     from which this clone was derived, please visit our web
                     site at  ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT       56 a     81 c     77 g     86 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 148;  Length 300;
  Best Local Similarity   80.0%;  Pred. No. 7.5e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||  ||| ||||| ||||||| |||
Db     27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  25
AI724043/c
LOCUS       AI724043      332 bp    mRNA            EST       19-JUL-2000
DEFINITION  RHIZ1_25_E10.y1_A001 Rhizome1 (RHIZ1) Sorghum halepense cDNA, mRNA
            sequence.
ACCESSION   AI724043
VERSION     AI724043.1  GI:5042895
KEYWORDS    EST.
SOURCE      Sorghum halepense.
  ORGANISM  Sorghum halepense
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Pratt,L.H. and Paterson,A.
  TITLE     An EST database from Sorghum: Sorghum halepense rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 4
            High quality sequence stop: 332
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .332
```

```
                     /organism="Sorghum halepense"
                     /db_xref="taxon:4560"
                     /clone_lib="Rhizome1 (RHIZ1)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       74 a     93 c     75 g     90 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 24;  Length 332;
  Best Local Similarity   80.0%;  Pred. No. 7.5e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    131 GCATGTCGATCCTCCTCTGGAGCCG 107


RESULT  26
AZ817869
LOCUS       AZ817869      360 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0087N04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0087N04 R, DNA sequence.
ACCESSION   AZ817869
VERSION     AZ817869.1  GI:12987777
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0087  row: N  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0087N04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      102 a     79 c     95 g     83 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 250;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 7.6e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| || | |||||| |||| |||
Db    102 GCATATCCAGCCTCGTTAGGTTCAG 126


RESULT  27
AW747232/c
LOCUS       AW747232      421 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_66_E04.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW747232
VERSION     AW747232.1  GI:7660970
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence stop: 409
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       96 a    120 c    111 g     94 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 120;  Length 421;
  Best Local Similarity   80.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    244 GCATGTCGATCCTCCTCTGGAGCCG 220
```

```
RESULT  28
BF387098
LOCUS       BF387098      432 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbk-c-10-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbk-c-10-0-UI 3', mRNA sequence.
ACCESSION   BF387098
VERSION     BF387098.1  GI:11371922
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bbk-c-10-0-UI"
                     /clone_lib="UI-R-CA1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex,  corpus striatum, testis, and
                     hippocampus.  For a detailed description of the  library
                     from which this clone was derived, please visit our web
                     site at  ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT       76 a    107 c    111 g    138 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 148;  Length 432;
  Best Local Similarity   80.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||  ||| ||||| ||||||| |||
Db     27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  29
BG241150/c
LOCUS       BG241150      446 bp    mRNA            EST       15-FEB-2001
DEFINITION  OV1_38_H01.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG241150
VERSION     BG241150.1  GI:12776223
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 413
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      100 a    128 c    121 g     97 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 175;  Length 446;
  Best Local Similarity   80.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    300 GCATGTCGATCCTCCTCTGGAGCCG 276


RESULT  30
AI283774/c
LOCUS       AI283774      456 bp    mRNA            EST       28-JAN-1999
DEFINITION  qj75e06.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1865314 3'
            similar to gb:M59979 PROSTAGLANDIN G/H SYNTHASE 1 PRECURSOR (HUMAN
            );; mRNA sequence.
ACCESSION   AI283774
VERSION     AI283774.1  GI:3922007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 2267 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

```
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1865314"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo. "
BASE COUNT      124 a    128 c    110 g     94 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 18;  Length 456;
  Best Local Similarity   80.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| | |||| | |||| ||||||
Db    260 GCATCTGGATCATTGTCATGTGCAG 236


RESULT  31
BG240831/c
LOCUS       BG240831      463 bp    mRNA            EST       15-FEB-2001
DEFINITION  OV1_38_H01.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG240831
VERSION     BG240831.1  GI:12775904
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 2
            High quality sequence stop: 461
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      107 a    131 c    124 g    100 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 175;  Length 463;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    300 GCATGTCGATCCTCCTCTGGAGCCG 276


RESULT  32
BG047654/c
LOCUS       BG047654      471 bp    mRNA            EST       25-JAN-2001
DEFINITION  OV1_26_E07.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG047654
VERSION     BG047654.1  GI:12497673
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 467
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      108 a    134 c    125 g    104 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 173;  Length 471;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    300 GCATGTCGATCCTCCTCTGGAGCCG 276


RESULT  33
```

```
BG048007/c
LOCUS       BG048007      473 bp    mRNA            EST       25-JAN-2001
DEFINITION  OV1_26_E07.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG048007
VERSION     BG048007.1  GI:12498325
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence stop: 465
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      109 a    134 c    127 g    103 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 173;  Length 473;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||||||||| || || || |
Db     300 GCATGTCGATCCTCCTCTGGAGCCG 276


RESULT  34
BF388049
LOCUS       BF388049      476 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbx-f-10-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbx-f-10-0-UI 3', mRNA sequence.
ACCESSION   BF388049
VERSION     BF388049.1  GI:11372873
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bbx-f-10-0-UI"
                     /clone_lib="UI-R-CA1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex,  corpus striatum, testis, and
                     hippocampus.  For a detailed description of the  library
                     from which this clone was derived, please visit our web
                     site at  ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      81 a    117 c    125 g    150 t      3 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 148;  Length 476;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||  ||| ||||| ||||||| |||
Db      27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  35
AW923923/c
LOCUS       AW923923      489 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_30_A09.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW923923
VERSION     AW923923.1  GI:8089748
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
```

```
          High quality sequence stop: 472
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      120 a    136 c    125 g    108 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 122;  Length 489;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| |||||||||| || || || |
Db    343 GCATGTCGATCCTCCTCTGGAGCCG 319


RESULT  36
BE355471/c
LOCUS       BE355471      497 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_114_H10.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355471
VERSION     BE355471.1  GI:9296573
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 2
            High quality sequence stop: 475
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      125 a    137 c    123 g    111 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 166;  Length 497;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| |||||||||| || || || |
Db    335 GCATGTCGATCCTCCTCTGGAGCCG 311


RESULT  37
AW747172/c
LOCUS       AW747172      502 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_66_E04.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW747172
VERSION     AW747172.1  GI:7660910
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 458
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      132 a    136 c    130 g    104 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 120;  Length 502;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| |||||||||| || || || |
Db    304 GCATGTCGATCCTCCTCTGGAGCCG 280


RESULT  38
AZ475043/c
LOCUS       AZ475043      506 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0292H17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0292H17 R, DNA sequence.
ACCESSION   AZ475043
VERSION     AZ475043.1  GI:10633168
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 506)
```

```
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0292  row: H  column: 17
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 506.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0292H17"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      176 a    106 c     92 g    132 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 244;  Length 506;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           | |||| ||| || | |||||||||
Db     427 GGATTTTGATACTTGCCAGGTGCAG 403


RESULT  39
BF390169
LOCUS       BF390169      510 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bcl-c-02-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bcl-c-02-0-UI 3', mRNA sequence.
ACCESSION   BF390169
VERSION     BF390169.1  GI:11375013
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bcl-c-02-0-UI"
                     /clone_lib="UI-R-CA1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex,  corpus striatum, testis, and
                     hippocampus.  For a detailed description of the  library
                     from which this clone was derived, please visit our web
                     site at  ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      90 a    130 c    131 g    158 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 148;  Length 510;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||  ||| ||||| |||||||| |||
Db      27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  40
BE364076/c
LOCUS       BE364076      511 bp    mRNA            EST       20-JUL-2000
DEFINITION  PI1_11_A09.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE364076
VERSION     BE364076.1  GI:9305633
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
```

```
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 503
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .511
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Pathogen induced 1 (PI1)"
                      /note="Organ: Anthracnose-infected leaves from
                      two-week-old sorghum plants 48 hr after inoculation;
                      Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                      Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                      cultivar) were infected with pathogen (isolate FRM42I of
                      Colletotrichum graminicola, which is a sorghum isolate).
                      RNA was prepared from infected leaves harvested from 45
                      seedlings  48 hours after inoculation.  Note: young
                      seedlings (2 weeks old) exhibit juvenile resistant
                      reaction, which is an incompatible interaction. As they
                      grow older (4 weeks or older), plants resume susceptiblity
                      to anthracnose disease. The library was made from poly-A
                      RNA in the cloning vector lambda ZAP II. Clones to be
                      sequenced were prepared by mass excision. WARNING: While
                      most or all ESTs are expected to derive from the host
                      plant, no effort was made to eliminate ESTs deriving from
                      the pathogen."
BASE COUNT      108 a    138 c    116 g    147 t      2 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 166;  Length 511;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    191 GCATGTCGATCCTCCTCTGGAGCCG 167


RESULT  41
BE364141/c
LOCUS       BE364141      511 bp    mRNA            EST       20-JUL-2000
DEFINITION  PI1_11_A09.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE364141
VERSION     BE364141.1  GI:9305698
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence stop: 511
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .511
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Pathogen induced 1 (PI1)"
                      /note="Organ: Anthracnose-infected leaves from
                      two-week-old sorghum plants 48 hr after inoculation;
                      Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                      Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                      cultivar) were infected with pathogen (isolate FRM42I of
                      Colletotrichum graminicola, which is a sorghum isolate).
                      RNA was prepared from infected leaves harvested from 45
                      seedlings  48 hours after inoculation.  Note: young
                      seedlings (2 weeks old) exhibit juvenile resistant
                      reaction, which is an incompatible interaction. As they
                      grow older (4 weeks or older), plants resume susceptiblity
                      to anthracnose disease. The library was made from poly-A
                      RNA in the cloning vector lambda ZAP II. Clones to be
                      sequenced were prepared by mass excision. WARNING: While
                      most or all ESTs are expected to derive from the host
                      plant, no effort was made to eliminate ESTs deriving from
                      the pathogen."
BASE COUNT      108 a    138 c    116 g    149 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 166;  Length 511;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||||||||| || || || |
Db    191 GCATGTCGATCCTCCTCTGGAGCCG 167


RESULT  42
AQ940100
LOCUS       AQ940100      515 bp    DNA             GSS       27-JAN-2000
DEFINITION  Sheared DNA-43C24.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-43C24, DNA sequence.
ACCESSION   AQ940100
VERSION     AQ940100.1  GI:6763365
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-43C24.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .515
```

```
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-43C24"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
BASE COUNT      106 a    158 c    119 g    132 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 235;  Length 515;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          || |||||||||  | ||  ||||||
Db     99 GCCTTTCGATCCGCCTCCCGTGCAG 123


RESULT  43
BE355404/c
LOCUS       BE355404      518 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_114_H10.b2_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355404
VERSION     BE355404.1  GI:9296506
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 511
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      130 a    149 c    126 g    113 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 166;  Length 518;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| |||||||||  || ||  || |
Db    351 GCATGTCGATCCTCCTCTGGAGCCG 327


RESULT  44
BE355326/c
LOCUS       BE355326      519 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_114_H10.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355326
VERSION     BE355326.1  GI:9296428
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 512
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      130 a    149 c    127 g    113 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 166;  Length 519;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| |||||||||  || ||  || |
Db    351 GCATGTCGATCCTCCTCTGGAGCCG 327


RESULT  45
TA367G09Q
LOCUS       TA367G09Q     529 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 367g09, reverse sequence,
            genomic survey sequence.
ACCESSION   AL495254
VERSION     AL495254.1  GI:11871641
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
```

```
            Trypanosoma.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="367g09"
BASE COUNT      117 a    167 c    114 g    131 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 258;  Length 529;
  Best Local Similarity   80.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           || ||||||||| | ||  ||||||
Db     455 GCCTTTCGATCCGCCTCCCGTGCAG 479


Search completed: October  9, 2001, 13:46:49
Job time: 9524 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:14:30 ; Search time 1666.31 Seconds
(without alignments)
232.066 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 BioB gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

	9	25	100.0	5872	9	A93679
	10	25	100.0	5872	9	AR101809
	11	25	100.0	5872	9	AR101810
	12	25	100.0	11022	1	AE000180
	13	25	100.0	13501	1	AE005258
	14	25	100.0	297816	2	AP002553
	15	23.4	93.6	5526	2	AF250776
	16	21.8	87.2	7215	3	SMABIO
	17	20.8	83.2	4000	2	AF250768
	18	20.2	80.8	1041	3	EHBIOTOP2
c	19	18.2	72.8	160	97	HSY15922
	20	18.2	72.8	525	7	CEU10160
	21	18.2	72.8	987	9	AR008950
	22	18.2	72.8	987	10	I28591
	23	18.2	72.8	993	7	CEU57752
	24	18.2	72.8	1005	91	CTU19834
	25	18.2	72.8	1007	89	AF180563
	26	18.2	72.8	1012	7	AF007576
	27	18.2	72.8	1012	7	BTU11815
	28	18.2	72.8	1018	9	AR052857
	29	18.2	72.8	1058	7	ECIL12P40
	30	18.2	72.8	1080	97	MMU19841
	31	18.2	72.8	1389	10	AX084106
	32	18.2	72.8	1399	9	AR091391
	33	18.2	72.8	1399	97	HUMCLMF40
	34	18.2	72.8	1560	9	AR091395
	35	18.2	72.8	1623	9	AR091394
	36	18.2	72.8	1870	9	A92079
	37	18.2	72.8	2318	97	HUMNKSFP40
	38	18.2	72.8	2362	9	AR082684
	39	18.2	72.8	2362	10	I28325
	40	18.2	72.8	2362	10	I89770
c	41	18.2	72.8	2611	93	HSAJ2967
	42	18.2	72.8	6139	9	AR091393
	43	18.2	72.8	18340	91	AY008847
	44	18.2	72.8	33692	61	AC010294
	45	18.2	72.8	42729	3	MTCY71

A93679 Sequence 6
AR101809 Sequence
AR101810 Sequence
AE000180 Escherich
AE005258 Escherich
AP002553 Escherich
AF250776 Unculture
D17468 S. marcesce
AF250768 Unculture
U38648 Erwinia her
Y15922 Homo sapien
U10160 Cervus elap
AR008950 Sequence
I28591 Sequence 1
U57752 Cervus elap
U19834 Cercocebus
AF180563 Homo sapi
AF007576 Capra hir
U11815 Bos taurus
AR052857 Sequence
Y11129 Equus cabal
U19841 Macaca mula
AX084106 Sequence
AR091391 Sequence
M65272 Human cytot
AR091395 Sequence
AR091394 Sequence
A92079 Sequence 3
M65290 Human natur
AR082684 Sequence
I28325 Sequence 1
I89770 Sequence 1
AJ222967 Homo sapi
AR091393 Sequence
AY008847 Homo sapi
AC010294 Homo sapi
Z92771 Mycobacteri

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.27;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   2
```

```
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.27;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.26;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
```

```
Db    128 CGATCCTCGTCAGGTGCAGGTCAGC 152


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 10;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    146 CGATCCTCGTCAGGTGCAGGTCAGC 170


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match           100.0%;  Score 25;  DB 2;  Length 5793;
  Best Local Similarity  100.0%;  Pred. No. 0.21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db   2116 CGATCCTCGTCAGGTGCAGGTCAGC 2140


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
```

```
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a    1552 c    1695 g    1307 t

ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
```

YINLAFVDASTLGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcagc 25
||||||||||||||||||||||||||
Db 221 CGATCCTCGTCAGGTGCAGGTCAGC 245

RESULT 10
AR101809
LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcagc 25
||||||||||||||||||||||||||
Db 221 CGATCCTCGTCAGGTGCAGGTCAGC 245

RESULT 11
AR101810
LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcagc 25
||||||||||||||||||||||||||
Db 221 CGATCCTCGTCAGGTGCAGGTCAGC 245

RESULT 12
AE000180
LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"

```
                /db_xref="taxon:83333"
repeat_region   <1. .130
                /note="REP (repetitive extragenic palindromic) element;
                contains 4 REP sequences"
gene            complement(147. .1430)
                /gene="ybhC"
                /note="b0772"
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL

  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db    3597 CGATCCTCGTCAGGTGCAGGTCAGC 3621


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
```

```
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
```

```
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db    9683 CGATCCTCGTCAGGTGCAGGTCAGC 9707


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
```

```
gene            1245. .2114
                /gene="ECs0754"
CDS             1245. .2114
                /gene="ECs0754"
                /note="similar to SUCD_ECOLI gi|1786949 percent identity
                100 in 289 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="succinyl-CoA synthetase alpha subunit"
                /protein_id="BAB34177.1"
                /db_xref="GI:13360213"
                /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
gene            2781. .3686
                /gene="ECs0755"
CDS             2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db  92415 CGATCCTCGTCAGGTGCAGGTCAGC 92439
```

```
RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
```

```
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             93.6%;  Score 23.4;  DB 2;  Length 5526;
  Best Local Similarity   96.0%;  Pred. No. 1.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           ||||||||||||||| |||||||||
Db    2067 CGATCCTCGTCAGGTCCAGGTCAGC 2091


RESULT  16
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
  JOURNAL   Unpublished (1993)
```

```
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
            Phone: 048-433-2545
            Email: nsakurai@ddbj.nig.ac.jp
            Fax:   048-433-2540.
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a   2225 c   2367 g   1276 t
ORIGIN


  Query Match             87.2%;  Score 21.8;  DB 3;  Length 7215;
  Best Local Similarity   92.0%;  Pred. No. 6.6;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          ||| || ||||||||||||||||||||
Db   2823 CGACCCGCGTCAGGTGCAGGTCAGC 2847


RESULT  17
AF250768
LOCUS       AF250768     4000 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
            cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
            complete cds; and biotin biosynthesis protein BioC (bioC) gene,
            partial cds.
ACCESSION   AF250768
VERSION     AF250768.1  GI:12620095
KEYWORDS    .
SOURCE      uncultured bacterium pCosFS1.
  ORGANISM  uncultured bacterium pCosFS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .4000
                     /organism="uncultured bacterium pCosFS1"
                     /db_xref="taxon:143795"
                     /clone="pCosFS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .1177)
                     /gene="bioA"
     CDS             complement(<1. .1177)
```

```
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="BioA-like protein"
                     /protein_id="AAG60558.1"
                     /db_xref="GI:12620098"
                     /translation="MSVTDSDLEFDQRHIWHPYTSMSHPLPCYPVEAASGVELQLADG
                     RRLIDGMSSWWAAIHGYNHPQLNLAATQQLEKMSHVMFGGITHPAANELCRKVVAMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCGFDQAWDENDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLERVRELCDRYQVLLIADEIATGFGRTGKLFACEHAEVVPDILCLGK
                     ALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVANASLSLLAENRW
                     QQQVAAIETQLKQQLLPLKMQPKVADVRVLGAIGVVEMREPVDMAQLQRGFVERGVWI
                     "
     gene            1265. .2305
                     /gene="bioB"
     CDS             1265. .2305
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="BioB-like protein"
                     /protein_id="AAG60559.1"
                     /db_xref="GI:12620099"
                     /translation="MMADRIHWTVGLAQTLFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSSRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGSQAERLAEAGLDYYNHNLDTSPE
                     FYGNIITTRSYQERLDTLGKVRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTETEHGDNQQQEVLAARLM
                     QADTAEFYNAAL"
     gene            2305. .3456
                     /gene="bioF"
     CDS             2305. .3456
                     /gene="bioF"
                     /note="7-keto-8-aminopelargonic acid synthetase; similar
                     to BioF"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA-synthetase"
                     /protein_id="AAG60556.1"
                     /db_xref="GI:12620096"
                     /translation="MSWQQRIEQALEERXQSAAYRHRLVIKGGNGRYLQLNDQRYLNF
                     SGNDYLGLSQNAQVIAAWQQGAQQYGVGSGGSGHVTGFCTAHQALEQQLAQWLGYPRS
                     LLFISGYAANQALLAALMQGEDRILADRLSHASLLEAAAHSPAQLRRFRHNQPQALAD
                     LLAKPCSGQTLAVTEGVFSMDGDSAPLAELYRLTRDAGAWLMVDDAHGIGVQGEQGRG
                     SCWQQGVRPELLVVTFGKAFGLSGAAVLCDESTAEYLLQFARHLIYSTAMPPAQACAL
                     QAALHCIQQGDELRTRLNTNIQRFRQGAAQLSLTLTESLTAIQPLLVGDNQRALDLAQ
                     HLRGQGLWVSAIRPPTVPPGGARLRITLTAAHQPQDIDRLLEVLHDVSV"
     gene            3440. .>4000
                     /gene="bioC"
     CDS             3440. .>4000
                     /gene="bioC"
                     /note="similar to BioC; probably reaction step prior to
                     synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60557.1"
                     /db_xref="GI:12620097"
                     /translation="MTSAFEQVNKQAVASAFSRAAVGYDAAAVLQRDVGERLLGMGQE
                     HQGEQLLDAGCGTGYFSRRWRELGKQVTALDLAPGMLAFARQQQAADHYLLGDIENVP
                     LGDAAVDISFSSLVVQWCSDLPRALAELYRVTRPGGVILFSTLAEGSLHELGDAWQQV
                     DGERHVNDFLPLAQIEAACAGYRHHLQ"
BASE COUNT      794 a   1164 c   1236 g    802 t      4 others
ORIGIN


  Query Match             83.2%;  Score 20.8;  DB 2;  Length 4000;
  Best Local Similarity   91.7%;  Pred. No. 21;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||| |||||||||||||| |||

Db   1373 GATCCGCGTCAGGTGCAGGTGAGC 1396


RESULT  18
EHBIOTOP2
LOCUS       EHBIOTOP2    1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             80.8%;  Score 20.2;  DB 3;  Length 1041;
  Best Local Similarity   88.0%;  Pred. No. 48;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          ||| ||||| ||||| |||||||||
Db    105 CGACCCTCGCCAGGTTCAGGTCAGC 129


RESULT  19
HSY15922/c
LOCUS       HSY15922      160 bp    DNA             PRI       06-APR-1998
DEFINITION  Homo sapiens CTNS gene, exon 1 and 3' flanking intronic region.
ACCESSION   Y15922
VERSION     Y15922.1  GI:3036848
KEYWORDS    CTNS gene; cystinosin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160)
  AUTHORS   Antignac,C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1997) Antignac C., U423, INSERM, Tour Lavoisier
            Hopital Necker 149 rue de Sevres Paris, 75015, FRANCE
REFERENCE   2  (bases 1 to 160)
  AUTHORS   Town,M., Jean,G., Cherqui,S., Attard,M., Forestier,L.,
            Whitmore,S.A., Callen,D.F., Gribouval,O., Broyer,M., Bates,G.P.,
            van't Hoff,W. and Antignac,C.
  TITLE     A novel gene encoding an integral membrane protein is mutated in
            nephropathic cystinosis
  JOURNAL   Nat. Genet. 18 (4), 319-324 (1998)
  MEDLINE   98196758
FEATURES             Location/Qualifiers
     source          1. .160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     exon            1. .110
                     /gene="CTNS"
                     /number=1
                     /usedin=Y15924:CTNSmRNA
     gene            1. .110
                     /gene="CTNS"
     intron          111. .>160
                     /number=1
BASE COUNT       22 a     55 c     57 g     26 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 97;  Length 160;
  Best Local Similarity   87.0%;  Pred. No. 5.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      67 GAGCCTCGCCAGGTGCCGGTCAG 45


RESULT  20
CEU10160
LOCUS       CEU10160      525 bp    mRNA            MAM       09-JUL-1994
DEFINITION  Cervus elaphus interleukin 12 mRNA, partial cds.
ACCESSION   U10160
VERSION     U10160.1  GI:508867
KEYWORDS    .
SOURCE      red deer.
  ORGANISM  Cervus elaphus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
            Cervidae; Cervinae; Cervus.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Lockhart,E.A.
  TITLE     Cloning and sequencing of cervine IL-12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 525)
  AUTHORS   Lockhart,E.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1994) Euan A. Lockhart, Microbiology, Deer
            Research Laboratory, University of Otago, Union St., Dunedin, New
            Zealand
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Cervus elaphus"
                     /db_xref="taxon:9860"
     CDS             <1. .>525
                     /standard_name="IL-12"
                     /codon_start=2
                     /product="interleukin 12"
                     /protein_id="AAA19658.1"
                     /db_xref="GI:508868"
                     /translation="AGQYTCHKGGEVLSRSLLLLHKKEDGIWSTDILKDQKEPKAKSF
                     LKCEAKDYSGHFTCWWLTAISTDLKFSVKSSRGSSDPRGVTCGAMSLSTEKVIVDHRE
                     YKKYTVECQEGSACPAAEESLPIEVVVEAVHKLKYENYTSSFFIRDIIKPDPPKNLQL
                     RPLKNSRQVEVSWE"
BASE COUNT      154 a    120 c    142 g    109 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 7;  Length 525;
  Best Local Similarity   87.0%;  Pred. No. 4.5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     495 ATTCTCGGCAGGTGGAGGTCAGC 517


RESULT  21
AR008950
LOCUS       AR008950      987 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 1 from patent US 5756085.
ACCESSION   AR008950
VERSION     AR008950.1  GI:3967755
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Sykes,M. and Wolf,S.F.
  TITLE     Use of interleukin-12 to prevent graft versus host disease
  JOURNAL   Patent: US 5756085-A 1 26-MAY-1998;
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="unknown"
BASE COUNT      264 a    243 c    259 g    221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  22
I28591
LOCUS       I28591        987 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 1 from patent US 5573764.
ACCESSION   I28591
VERSION     I28591.1  GI:1819367
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Sykes,M. and Wolf,S.F.
  TITLE     Use of interleukin-12 to prevent graft versus host disease
  JOURNAL   Patent: US 5573764-A 1 12-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="unknown"
BASE COUNT      264 a    243 c    259 g    221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  23
```

```
CEU57752
LOCUS       CEU57752      993 bp    mRNA            MAM       28-OCT-1999
DEFINITION  Cervus elaphus interleukin-12 p40 subunit mRNA, complete cds.
ACCESSION   U57752
VERSION     U57752.1  GI:1373387
KEYWORDS    .
SOURCE      red deer.
  ORGANISM  Cervus elaphus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
            Cervidae; Cervinae; Cervus.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Lockhart,E., Slobbe,L. and Buchan,G.
  TITLE     The characterisation of a cervine immunoregulatory cytokine,
            interleukin 12
  JOURNAL   DNA Seq. 10 (2), 139-148 (1999)
  MEDLINE   99304506
   PUBMED   10376218
REFERENCE   2  (bases 1 to 993)
  AUTHORS   Lockhart,E.A. and Buchan,G.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1996) Euan A. Lockhart, Microbiology, University
            of Otago, Cumberland St, Dunedin, New Zealand
FEATURES             Location/Qualifiers
     source          1. .993
                     /organism="Cervus elaphus"
                     /db_xref="taxon:9860"
                     /cell_type="peripheral blood mononuclear cells"
     primer_bind     1. .33
     CDS             1. .984
                     /codon_start=1
                     /product="interleukin-12 p40 subunit"
                     /protein_id="AAB02258.1"
                     /db_xref="GI:1373388"
                     /translation="MHPQQLVVSWFSLVLLTSPIVAIWELEKNVYVVELDWYPDAPGE
                     TVVLRCDTPEEDGITWTSDQSSEVLGSGKTLTVQVKEFGDAGQYTCHKGGEVLSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAASLSTEKVIVDHREYKKYTVECQEGSACPAAEESLPIEVVVEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFMDQTSAKVTCHKDASIRVQARDRYYNSFWSEWASVSCS"
     primer_bind     complement(961. .993)
BASE COUNT      269 a    240 c    263 g    221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 7;  Length 993;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  24
CTU19834
LOCUS       CTU19834     1005 bp    mRNA            PRI       03-FEB-1996
DEFINITION  Cercocebus torquatus interleukin-12 alpha p40 subunit (IL-12a)
            mRNA, complete cds.
ACCESSION   U19834
VERSION     U19834.1  GI:644785
KEYWORDS    .
SOURCE      red-crowned mangabey.
  ORGANISM  Cercocebus torquatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Cercocebus.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Villinger,F., Brar,S.S., Mayne,A., Chikkala,N. and Ansari,A.A.
  TITLE     Comparative sequence analysis of cytokine genes from human and
            nonhuman primates
  JOURNAL   J. Immunol. 155 (8), 3946-3954 (1995)
  MEDLINE   96003435
REFERENCE   2  (bases 1 to 1005)
  AUTHORS   Villinger,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1995) Francois J. Villinger, Emory University,
            Pathology, 1327 Clifton Rd, Atlanta, GA 30322, USA
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Cercocebus torquatus"
                     /sub_species="atys"
                     /db_xref="taxon:9530"
                     /clone="FPn IL-12a"
                     /cell_type="peripheral blood mononuclear cells"
                     /tissue_type="blood"
     gene            9. .995
                     /gene="IL-12a"
     CDS             9. .995
                     /gene="IL-12a"
                     /codon_start=1
                     /function="NK cell stimulatory factor"
                     /product="interleukin-12 alpha p40 subunit"
                     /protein_id="AAA86701.1"
                     /db_xref="GI:644786"
                     /translation="MCHQQLVISWFSLVFLASPLMAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSGEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSHSLL
                     LPHKKEDGIWSTDILKDQKEPKNETFLRCEAKNYSGRITCWWLSTISTDLTFSIISSR
                     GSSNPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEERLPIEVMVDAIHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCIQ
                     VQGKSKREKKDRIFTDKTSATVICRKNASFSVQAQDRYYSSSWNEWTSVPCS"
BASE COUNT      275 a    252 c    257 g    221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 91;  Length 1005;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    751 ATTCTCGGCAGGTGGAGGTCAGC 773


RESULT  25
AF180563
LOCUS       AF180563     1007 bp    mRNA            PRI       26-SEP-1999
DEFINITION  Homo sapiens interleukin 12, P40 mRNA, complete cds.
ACCESSION   AF180563
VERSION     AF180563.1  GI:5923854
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   Hongyuan,J. and Meiyun,Z.
  TITLE     Cloning and sequence analysis of IL-12 cDNA from Chinese
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1007)
  AUTHORS   Hongyuan,J. and Meiyun,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) Hepatitis, Institute of Virology, CAPM, 100
            Yingxinjie, Xuanwuqu, Beijing 100052, China
FEATURES             Location/Qualifiers
     source          1. .1007
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="peripheral mononuclear"
     CDS             7. .993
                     /note="NKSF/CLMF; IL12p40; 40kD"
                     /codon_start=1
                     /product="interleukin 12, P40"
                     /protein_id="AAD56386.1"
                     /db_xref="GI:5923855"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
```

```
                    MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                    LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                    GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                    KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                    VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
BASE COUNT      270 a    247 c    265 g    225 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 89;  Length 1007;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    749 ATTCTCGGCAGGTGGAGGTCAGC 771


RESULT  26
AF007576
LOCUS       AF007576     1012 bp    mRNA            MAM       14-JUL-1997
DEFINITION  Capra hircus interleukin 12 p40 subunit (IL-12) mRNA, complete cds.
ACCESSION   AF007576
VERSION     AF007576.1  GI:2253433
KEYWORDS    .
SOURCE      goat.
  ORGANISM  Capra hircus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Capra.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   Beyer,J.C., Cheevers,W.P. and Kumpula-McWhirter,N.M.
  TITLE     CLoning of caprine IL-12 40 kDa subunit
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1012)
  AUTHORS   Beyer,J.C., Cheevers,W.P. and Kumpula-McWhirter,N.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1997) Veterinary Microbiology and Pathology,
            Washington State University, P.O. Box 647040, Pullman, WA
            99164-7040, USA
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Capra hircus"
                     /db_xref="taxon:9925"
                     /cell_type="LPS-stimulated microglial cell"
     gene            1. .1012
                     /gene="IL-12"
     CDS             1. .984
                     /gene="IL-12"
                     /codon_start=1
                     /product="interleukin 12 p40 subunit"
                     /protein_id="AAB62945.1"
                     /db_xref="GI:2253434"
                     /translation="MHPQQLVVSWFSLVLLASPIVAIWELEKNVYVVELDWYPNAPGE
                     TVVLTCDTPEEDGITWTSDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCSWLTAISTNLKFSVKSSR
                     GSSDPRGVTCGAASLSAEKVSMDHREYNKYTVECQEGSACPAAEESLPIEVVMEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFTDQTSAKVTCHKDANIRVQARDRYYSSFWSEWASVSCS"
BASE COUNT      280 a    253 c    260 g    219 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 7;  Length 1012;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  27
BTU11815
LOCUS       BTU11815     1012 bp    mRNA            MAM       27-JAN-1996
DEFINITION  Bos taurus interleukin 12 40 kDa subunit mRNA, complete cds.
ACCESSION   U11815
VERSION     U11815.1  GI:555795
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   Zarlenga,D.S., Canals,A., Aschenbrenner,R.A. and Gasbarre,L.C.
  TITLE     Enzymatic amplification and molecular cloning of cDNA encoding the
            small and large subunits of bovine interleukin 12
  JOURNAL   Biochim. Biophys. Acta 1270 (2-3), 215-217 (1995)
  MEDLINE   95244628
REFERENCE   2  (bases 1 to 1012)
  AUTHORS   Zarlenga,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1994) Dante S. Zarlenga, U.S. Department of
            Agriculture, ARS, LPSI, PIL, Beltsville, MD 20705, USA
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Bos taurus"
                     /strain="Holstein"
                     /db_xref="taxon:9913"
                     /clone="BIL12 40"
                     /sex="male"
                     /cell_type="lymphocytes and macrophages"
                     /tissue_type="lymphoid"
     CDS             1. .984
                     /codon_start=1
                     /product="interleukin 12 40 kDa subunit"
                     /protein_id="AAA85792.1"
                     /db_xref="GI:555796"
                     /translation="MHPQQLVVSWFSLVLLASPIVAMWELEKNVYVVELDWYPDAPGE
                     TVVLTCDTPEEDGITWTSDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAALLSAEKVSLEHREYNKYTVECQEGSACPAAEESLLIEVVVEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFMDQTSAKVTCHKDANVRVQARDRYYSSFWSEWASVSCS"
BASE COUNT      275 a    245 c    267 g    225 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 7;  Length 1012;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  28
AR052857
LOCUS       AR052857     1018 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 194 from patent US 5833975.
ACCESSION   AR052857
VERSION     AR052857.1  GI:5977719
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1018)
  AUTHORS   Paoletti,E., Tartaglia,J. and Cox,W.I.
  TITLE     Canarypox virus expressing cytokine and/or tumor-associated antigen
            DNA sequence
  JOURNAL   Patent: US 5833975-A 194 10-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .1018
                     /organism="unknown"
```

```
BASE COUNT      283 a    245 c    260 g    230 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT  29
ECIL12P40
LOCUS       ECIL12P40    1058 bp    mRNA            MAM       17-NOV-1999
DEFINITION  Equus caballus mRNA for interleukin 12 p40 subunit.
ACCESSION   Y11129
VERSION     Y11129.1  GI:5441622
KEYWORDS    interleukin-12 p40.
SOURCE      horse.
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Nicolson,L., Penha-Goncalves,M.N., Keanie,J.L., Logan,N.A.,
            Argyle,D.J. and Onions,D.E.
  TITLE     Cloning and sequencing of horse interleukin-12 and interleukin-18
            cDNAs
  JOURNAL   Immunogenetics 50 (1-2), 94-97 (1999)
  MEDLINE   20009534
REFERENCE   2  (bases 1 to 1058)
  AUTHORS   Nicolson,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1997) L. Nicolson, Dept. of Vet. Pathology,
            University of Glasgow Vet. School, Bearsden Rd., Glasgow G61 1QH,
            UK
FEATURES             Location/Qualifiers
     source          1. .1058
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /cell_type="alveolar macrophages"
                     /cell_line="LPS-stimulated alveolar macrophages"
     gene            1. .990
                     /gene="IL-12 p40"
     misc_feature    1. .12
                     /gene="IL-12 p40"
                     /note="primer-derived sequence"
     CDS             1. .990
                     /gene="IL-12 p40"
                     /codon_start=1
                     /product="interleukin 12 p40 subunit"
                     /protein_id="CAA72011.1"
                     /db_xref="GI:5441623"
                     /translation="MCHQWLVLSWFSLVLLASPLMAIWELEKDVYVVELDWYPDAPGE
                     MVVLTCNTPEEEGITWTSAQSNEVLGSGKTLTIQVKEFGDAGWYTCHKGGEVLSHSHL
                     LLHKKEDGIWSTDILKDQKESKNKTFLKCEAKNYSGRFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAATLSAERVSVDDREYKKYTVECQEGSACPAAEESLPIEIVVDAVHK
                     LKYENYTSGFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPETWSTPHSYFSLTFSI
                     QVQGKNKKERKDRLFMDETSATVTCHKDGQIRVQARDRYYSSSWSEWASVSCS"
     sig_peptide     1. .66
                     /gene="IL-12 p40"
     variation       714
                     /gene="IL-12 p40"
                     /note="allelic difference, no change in amino acid
                     sequence"
                     /replace="g"
     misc_feature    1041. .1058
                     /note="primer-derived sequence"
BASE COUNT      284 a    269 c    282 g    223 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 7;  Length 1058;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  30
MMU19841
LOCUS       MMU19841     1080 bp    mRNA            PRI       03-FEB-1996
DEFINITION  Macaca mulatta interleukin-12 alpha (IL-12a) mRNA, complete cds.
ACCESSION   U19841
VERSION     U19841.1  GI:644799
KEYWORDS    .
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Villinger,F., Brar,S.S., Mayne,A., Chikkala,N. and Ansari,A.A.
  TITLE     Comparative sequence analysis of cytokine genes from human and
            nonhuman primates
  JOURNAL   J. Immunol. 155 (8), 3946-3954 (1995)
  MEDLINE   96003435
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Villinger,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1995) Francois J. Villinger, Emory University,
            Pathology, 1327 Clifton Rd, Atlanta, GA 30322, USA
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /clone="RSo IL-12a clone 4"
                     /cell_type="peripheral blood mononuclear cells"
                     /tissue_type="blood"
     gene            13. .999
                     /gene="IL-12a"
     CDS             13. .999
                     /gene="IL-12a"
                     /codon_start=1
                     /function="NK cell stimulatory factor"
                     /product="interleukin-12 alpha"
                     /protein_id="AAB60397.1"
                     /db_xref="GI:644800"
                     /translation="MCHQQLVISWFSLVFLASPLMAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSGEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSHSLL
                     LLHKKEDGIWSTDVLKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSNPQGVTCGAVTLSAERVRGDNKEYEYSVECQEDSACPAAEERLPIEVMVDAIHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCIQ
                     VQGKSKREKKDRIFTDKTSATVICRKNASFSVQAQDRYYSSSWSEWASVPCS"
     variation       329
                     /gene="IL-12a"
                     /replace="t"
     variation       405
                     /gene="IL-12a"
                     /replace="c"
     variation       519
                     /gene="IL-12a"
                     /replace="a"
     variation       629
                     /gene="IL-12a"
                     /note="Leu to Pro"
                     /replace="c"
     variation       665
                     /gene="IL-12a"
                     /note="Leu to Pro"
                     /replace="c"
     variation       714
                     /gene="IL-12a"
```

```
                     /replace="g"
     variation       878
                     /gene="IL-12a"
                     /note="Ile to Thr"
                     /replace="c"
     variation       895
                     /gene="IL-12a"
                     /note="Ser to Pro"
                     /replace="c"
BASE COUNT      309 a    263 c    276 g    232 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 97;  Length 1080;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    755 ATTCTCGGCAGGTGGAGGTCAGC 777


RESULT  31
AX084106
LOCUS       AX084106     1389 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 4 from Patent WO0110912.
ACCESSION   AX084106
VERSION     AX084106.1  GI:13185656
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1389)
  AUTHORS   Gillies,S.D. and Lo,K.M.
  TITLE     Multiple cytokine-antibody complexes
  JOURNAL   Patent: WO 0110912-A 4 15-FEB-2001;
            Lexigen Pharmaceuticals Corp. (US)
FEATURES             Location/Qualifiers
     source          1. .1389
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="human p40-IL-2 fusion protein coding sequence"
BASE COUNT      413 a    325 c    327 g    324 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 1389;
  Best Local Similarity   87.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  32
AR091391
LOCUS       AR091391     1399 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 3 from patent US 5994104.
ACCESSION   AR091391
VERSION     AR091391.1  GI:10018146
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1399)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 3 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1399
                     /organism="unknown"
BASE COUNT      390 a    310 c    333 g    366 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1399;
  Best Local Similarity   87.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    785 ATTCTCGGCAGGTGGAGGTCAGC 807


RESULT  33
HUMCLMF40
LOCUS       HUMCLMF40    1399 bp    mRNA            PRI       27-APR-1993
DEFINITION  Human cytotoxic lymphocyte maturation factor 40 kDa subunit mRNA,
            complete cds.
ACCESSION   M65272 M38443 M38444
VERSION     M65272.1  GI:180625
KEYWORDS    cytotoxic lymphocyte maturation factor 40 kDa subunit; interleukin.
SOURCE      Human, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1399)
  AUTHORS   Gubler,U., Chua,A.O., Schoenhaut,D.S., Dwyer,C.M., McComas,W.,
            Motyka,R., Nabavi,N., Wolitzky,A.G., Quinn,P.M., Familletti,P.C.
            and Gately,M.K.
  TITLE     Coexpression of two distinct genes is required to generate secreted
            bioactive cytotoxic lymphocyte maturation factor
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 4143-4147 (1991)
  MEDLINE   91239523
FEATURES             Location/Qualifiers
     source          1. .1399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             43. .1029
                     /codon_start=1
                     /product="cytotoxic lymphocyte maturation factor 40 kDa
                     subunit"
                     /protein_id="AAA35695.1"
                     /db_xref="GI:180626"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                     VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSSWSEWASVPCS"
BASE COUNT      390 a    310 c    333 g    366 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 97;  Length 1399;
  Best Local Similarity   87.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    785 ATTCTCGGCAGGTGGAGGTCAGC 807


RESULT  34
AR091395
LOCUS       AR091395     1560 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 11 from patent US 5994104.
ACCESSION   AR091395
VERSION     AR091395.1  GI:10018150
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
```

```
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 11 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1560
                     /organism="unknown"
BASE COUNT      443 a    375 c    393 g    349 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1560;
  Best Local Similarity   87.0%;  Pred. No. 3.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    680 ATTCTCGGCAGGTGGAGGTCAGC 702
```

```
RESULT  35
AR091394
LOCUS       AR091394     1623 bp    DNA              PAT       07-SEP-2000
DEFINITION  Sequence 9 from patent US 5994104.
ACCESSION   AR091394
VERSION     AR091394.1  GI:10018149
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1623)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 9 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1623
                     /organism="unknown"
BASE COUNT      449 a    393 c    408 g    373 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1623;
  Best Local Similarity   87.0%;  Pred. No. 3.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 ATTCTCGGCAGGTGGAGGTCAGC 765
```

```
RESULT  36
A92079
LOCUS       A92079       1870 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 3 from Patent WO9821322.
ACCESSION   A92079
VERSION     A92079.1  GI:6740883
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1870)
  AUTHORS   Junghans,C. and Wittig,B.
  TITLE     DESIGN PRINCIPLE FOR CONSTRUCTING EXPRESSION CONSTRUCTS FOR GENE
            THERAPY
  JOURNAL   Patent: WO 9821322-A 3 22-MAY-1998;
            JUNGHANS CLAAS (DE); SOFT GENE GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .1870
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      500 a    456 c    445 g    469 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1870;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db   1375 ATTCTCGGCAGGTGGAGGTCAGC 1397
```

```
RESULT  37
HUMNKSFP40
LOCUS       HUMNKSFP40   2318 bp    mRNA             PRI       07-JAN-1995
DEFINITION  Human natural killer cell stimulatory factor (NKSF) mRNA, complete
            cds, clone p40.
ACCESSION   M65290
VERSION     M65290.1  GI:189235
KEYWORDS    cytokine; natural killer cell stimulatory factor.
SOURCE      Human EBV-transformed B cell line RPMI 8866, cDNA to mRNA, clone
            p40.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2318)
  AUTHORS   Wolf,S.F., Temple,P.A., Kobayashi,M., Young,D., Dicig,M., Lowe,L.,
            Dzialo,R., Fitz,L., Ferenz,C., Hewick,R.M., Kelleher,K.,
            Herrmann,S.H., Clark,S.C., Azzoni,L., Chan,S.H., Trinchieri,G. and
            Perussia,B.
  TITLE     Cloning of cDNA for natural killer cell stimulatory factor, a
            heterodimeric cytokine with multiple biologic effects on T and
            natural killer cells
  JOURNAL   J. Immunol. 146 (9), 3074-3081 (1991)
  MEDLINE   91201875
FEATURES             Location/Qualifiers
     source          1. .2318
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="EBV-transformed B cell line RPMI 8866"
                     /map="Unassigned"
     sig_peptide     14. .79
                     /gene="NKSF2"
                     /note="G00-127-870"
     CDS             14. .1000
                     /gene="NKSF2"
                     /codon_start=1
                     /db_xref="GDB:G00-127-870"
                     /product="natural killer cell stimulatory factor"
                     /protein_id="AAA59938.1"
                     /db_xref="GI:189236"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPNNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                     VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
     gene            14. .1000
                     /gene="NKSF2"
     mat_peptide     80. .997
                     /gene="NKSF2"
                     /note="G00-127-870"
                     /product="natural killer cell stimulatory factor"
BASE COUNT      692 a    478 c    535 g    613 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 97;  Length 2318;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    756 ATTCTCGGCAGGTGGAGGTCAGC 778
```

```
RESULT  38
```

```
AR082684
LOCUS       AR082684      2362 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 1 from patent US 5976539.
ACCESSION   AR082684
VERSION     AR082684.1  GI:10009474
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5976539-A 1 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  39
I28325
LOCUS       I28325       2362 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 1 from patent US 5571515.
ACCESSION   I28325
VERSION     I28325.1  GI:1819101
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5571515-A 1 05-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  40
I89770
LOCUS       I89770       2362 bp    DNA             PAT       10-AUG-1998
DEFINITION  Sequence 1 from patent US 5723127.
ACCESSION   I89770
VERSION     I89770.1  GI:3409710
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5723127-A 1 03-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  41
HSAJ2967/c
LOCUS       HSAJ2967     2611 bp    mRNA            PRI       06-APR-1998
DEFINITION  Homo sapiens mRNA for cystinosin.
ACCESSION   AJ222967
VERSION     AJ222967.1  GI:3036839
KEYWORDS    cystinosin; CYSTN gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2611)
  AUTHORS   Antignac,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Antignac C., U423, INSERM, Tour Lavoisier
            Hopital Necker 149 rue de Sevres Paris, 75015, FRANCE
REFERENCE   2  (bases 1 to 2611)
  AUTHORS   Town,M., Jean,G., Cherqui,S., Attard,M., Forestier,L.,
            Whitmore,S.A., Callen,D.F., Gribouval,O., Broyer,M., Bates,G.P.,
            van't Hoff,W. and Antignac,C.
  TITLE     A novel gene encoding an integral membrane protein is mutated in
            nephropathic cystinosis
  JOURNAL   Nat. Genet. 18 (4), 319-324 (1998)
  MEDLINE   98196758
FEATURES             Location/Qualifiers
     source          1. .2611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            340. .1443
                     /gene="CTNS"
     CDS             340. .1443
                     /gene="CTNS"
                     /codon_start=1
                     /product="cystinosin"
                     /protein_id="CAA11021.1"
                     /db_xref="GI:3036840"
                     /db_xref="SPTREMBL:O60931"
                     /translation="MIRNWLTIFILFPLKLVEKCESSVSLTVPPVVKLENGSSTNVSL
                     TLRPPLNATLVITFEITFRSKNITILELPDEVVVPPGVTNSSFQVTSQNVGQLTVYLH
                     GNHSNQTGPRIRFLVIRSSAISIINQVIGWIYFVAWSISFYPQVIMNWRRKSVIGLSF
                     DFVALNLTGFVAYSVFNIGLLWVPYIKEQFLLKYPNGVNPVNSNDVFFSLHAVVLTLI
                     IIVQCCLYERGGQRVSWPAIGFLVLAWLFAFVTMIVAAVGVITWLQFLFCFSYIKLAV
                     TLVKYFPQAYMNFYYKSTEGWSIGNVLLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPT
                     KFGLGVFSIVFDVVFFIQHFCLYRKRPGYDQLN"
BASE COUNT      525 a    784 c    674 g    628 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 93;  Length 2611;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     67 GAGCCTCGCCAGGTGCCGGTCAG 45
```

```
RESULT  42
AR091393
LOCUS       AR091393     6139 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 7 from patent US 5994104.
ACCESSION   AR091393
VERSION     AR091393.1  GI:10018148
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6139)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 7 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .6139
                     /organism="unknown"
BASE COUNT     1552 a   1538 c   1558 g   1491 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 3.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    3964 ATTCTCGGCAGGTGGAGGTCAGC 3986


RESULT  43
AY008847
LOCUS       AY008847    18340 bp    DNA             PRI       17-NOV-2000
DEFINITION  Homo sapiens interleukin 12 p40 subunit (IL12B) gene, IL12B-1
            allele, complete cds.
ACCESSION   AY008847
VERSION     AY008847.1  GI:11192034
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 18340)
  AUTHORS   Huang,D., Cancilla,M.R. and Morahan,G.
  TITLE     Complete primary structure, chromosomal localization, and
            definition of polymorphisms of the gene encoding the human
            interleukin 12 p40 subunit
  JOURNAL   Genes Immun. 1 (8), 515-520 (2000)
REFERENCE   2  (bases 1 to 18340)
  AUTHORS   Huang,D., Cancilla,M.R. and Morahan,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2000) Genetics & Bioinformatics, The Walter and
            Eliza Hall Institute of Medical Research, PO Royal Melbourne
            Hospital, Victoria 3050, Australia
FEATURES             Location/Qualifiers
     source          1. .18340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q33"
     mRNA            join(2426. .2483,6133. .6220,9586. .9861,10404. .10521,
                     12395. .12609,14022. .14179,16100. .16231,16816. .18133)
                     /gene="IL12B"
                     /product="interleukin 12 p40 subunit"
     gene            2426. .18133
                     /gene="IL12B"
                     /note="IL12p40"
                     /allele="1"
     CDS             join(6133. .6220,9586. .9861,10404. .10521,12395. .12609,
                     14022. .14179,16100. .16231)
                     /gene="IL12B"
                     /note="cytokine"
                     /codon_start=1
                     /product="interleukin 12 p40 subunit"
                     /protein_id="AAG32620.1"
                     /db_xref="GI:11192035"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                     VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
BASE COUNT     5261 a   3884 c   3980 g   5215 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 91;  Length 18340;
  Best Local Similarity   87.0%;  Pred. No. 2.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db   14067 ATTCTCGGCAGGTGGAGGTCAGC 14089


RESULT  44
AC010294
LOCUS       AC010294    33692 bp    DNA             HTG       26-JAN-2000
DEFINITION  Homo sapiens chromosome 5 clone CTB-140P23, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC010294
VERSION     AC010294.1  GI:5882639
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 33692)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 33692)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------.
            * NOTE: This record contains 42 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      648: contig of 648 bp in length
            *                    gap of unknown length
            *      649     1296: contig of 648 bp in length
            *                    gap of unknown length
            *     1297     1926: contig of 630 bp in length
            *                    gap of unknown length
            *     1927     2816: contig of 890 bp in length
            *                    gap of unknown length
            *     2817     3812: contig of 996 bp in length
            *                    gap of unknown length
            *     3813     4002: contig of 190 bp in length
            *                    gap of unknown length
            *     4003     4249: contig of 247 bp in length
            *                    gap of unknown length
```

```
            *     4250     4677: contig of 428 bp in length
            *                  gap of unknown length
            *     4678     5300: contig of 623 bp in length
            *                  gap of unknown length
            *     5301     5958: contig of 658 bp in length
            *                  gap of unknown length
            *     5959     6598: contig of 640 bp in length
            *                  gap of unknown length
            *     6599     7183: contig of 585 bp in length
            *                  gap of unknown length
            *     7184     7338: contig of 155 bp in length
            *                  gap of unknown length
            *     7339     8020: contig of 682 bp in length
            *                  gap of unknown length
            *     8021     9055: contig of 1035 bp in length
            *                  gap of unknown length
            *     9056     9781: contig of 726 bp in length
            *                  gap of unknown length
            *     9782    10678: contig of 897 bp in length
            *                  gap of unknown length
            *    10679    11487: contig of 809 bp in length
            *                  gap of unknown length
            *    11488    12107: contig of 620 bp in length
            *                  gap of unknown length
            *    12108    12724: contig of 617 bp in length
            *                  gap of unknown length
            *    12725    12906: contig of 182 bp in length
            *                  gap of unknown length
            *    12907    13688: contig of 782 bp in length
            *                  gap of unknown length
            *    13689    14430: contig of 742 bp in length
            *                  gap of unknown length
            *    14431    15569: contig of 1139 bp in length
            *                  gap of unknown length
            *    15570    17161: contig of 1592 bp in length
            *                  gap of unknown length
            *    17162    17335: contig of 174 bp in length
            *                  gap of unknown length
            *    17336    18264: contig of 929 bp in length
            *                  gap of unknown length
            *    18265    19159: contig of 895 bp in length
            *                  gap of unknown length
            *    19160    20056: contig of 897 bp in length
            *                  gap of unknown length
            *    20057    21105: contig of 1049 bp in length
            *                  gap of unknown length
            *    21106    22217: contig of 1112 bp in length
            *                  gap of unknown length
            *    22218    22755: contig of 538 bp in length
            *                  gap of unknown length
            *    22756    23757: contig of 1002 bp in length
            *                  gap of unknown length
            *    23758    24309: contig of 552 bp in length
            *                  gap of unknown length
            *    24310    24642: contig of 333 bp in length
            *                  gap of unknown length
            *    24643    25964: contig of 1322 bp in length
            *                  gap of unknown length
            *    25965    27175: contig of 1211 bp in length
            *                  gap of unknown length
            *    27176    28598: contig of 1423 bp in length
            *                  gap of unknown length
            *    28599    29420: contig of 822 bp in length
            *                  gap of unknown length
            *    29421    30801: contig of 1381 bp in length
            *                  gap of unknown length
            *    30802    32587: contig of 1786 bp in length
            *                  gap of unknown length
            *    32588    33692: contig of 1105 bp in length.
FEATURES             Location/Qualifiers
     source          1. .33692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-140P23"
BASE COUNT     9394 a   7696 c   7805 g   8776 t     21 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 61;  Length 33692;
  Best Local Similarity   87.0%;  Pred. No. 2.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || |||| |||| ||||||||||
Db    7341 GACCCTCATCAGCTGCAGGTCAG 7363


RESULT  45
MTCY71
LOCUS       MTCY71      42729 bp    DNA             BCT       10-FEB-1999
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.
ACCESSION   Z92771 AL123456
VERSION     Z92771.1  GI:3242259
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 42729)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 42729)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 20, 1998 this sequence version replaced gi:2143309.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .42729
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
```

/clone="Y71"
misc_feature <1. .101
/note="fragment designated v015. Does not represent a physical clone"
gene 222. .1217
/gene="Rv3261"
CDS 222. .1217
/gene="Rv3261"
/note="Rv3261, (MTCY71.01, unknown), len: 331 aa, similar to U67566_3 Methanococcus jannaschii Q58653 hypothetical 34.6 kDa protein 1256 (311 aa)fasta scores, opt: 436, E(): 8.9e-22, (27.4% identity in 274 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3261"
/protein_id="CAB07094.1"
/db_xref="GI:1877316"
/db_xref="SPTREMBL:P96866"
/translation="MKVTVLAGGVGGARFLLGVQQLLGLGQFAANSAHSDADHQLSAV
VNVGDDAWIHGLRVCPDLDTCMYTLGGGVDPQRGWGQRDETWHAMQELVRYGVQPDWF
ELGDRDLATHLVRTQMLQAGYPLSQITEALCDRWQPGARLLPATDDRCETHVVITDPV
DESRKAIHFQEWWVRYRAQVPTHSFAFVGAEKSSAATEAIAALADADIIMLAPSNPVV
SIGAILAVPGIRAALREATAPIVGYSPIIGEKPLRGMADTCLSVIGVDSTAAAVGRHY
GARCATGILDCWLVHDGDHAEIDGVTVRSVPLLMTDPNATAEMVRAGCDLAGVVA"
gene 1214. .2560
/gene="Rv3262"
CDS 1214. .2560
/gene="Rv3262"
/note="Rv3262, (MTCY71.02, unknown), len: 448 aa, similar to AE000874_11 Methanobacterium thermoautotro O27098 CONSERVED PROTEIN (252 aa; E(): 7.6e-18; 37.3% identity in 249 aaoverlap). Also similar to Q58178 hypothetical 27.1 kDa protein 0768 (249 aa) opt: 375, E(): 6.9e-17, (33.9% identityin 239 aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3262"
/protein_id="CAB07089.1"
/db_xref="GI:1877317"
/db_xref="SPTREMBL:P96867"
/translation="MTGPEHGSASTIEILPVIGLPEFRPGDDLSAAVAAAAPWLRDGD
VVVVTSKVVSKCEGRLVPAPEDPEQRDRLRRKLIEDEAVRVLARKDRTLITENRLGLV
QAAAGVDGSNVGRSELALLPVDPDASAATLRAGLRERLGVTVAVVITDTMGRAWRNGQ
TDAAVGAAGLAVLRNYAGVRDPYGNELVVTEVAVADEIAAAADLVKGKLTATPVAVVR
GFGVSDDGSTARQLLRPGANDLFWLGTAEALELGRQQAQLLRRSVRRFSTDPVPGDLV
EAAVAEALTAPAPHHTRPTRFVWLQTPAIRARLLDRMKDKWRSDLTSDGLPADAIERR
VARGQILYDAPEVVIPMLVPDGAHSYPDAARTDAEHTMFTVAVGAAVQALLVALAVRG
LGSCWIGSTIFAADLVRDELDLPVDWEPLGAIAIGYADEPSGLRDPVPAADLLILK"
gene 2856. .4517
/gene="Rv3263"
CDS 2856. .4517
/gene="Rv3263"
/note="Rv3263, (MTCY71.03), len: 553, DNA methylase, (alternative start at aa 25), similar to eg MT57_ECOLI P25240 modification methylase eco57i (544 aa), fasta scores, opt: 595, E(): 0, (30.4% identity in 507 aa overlap); contains PS00092 N-6 Adenine-specific DNA methylases signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3263"
/protein_id="CAB07090.1"
/db_xref="GI:1877318"
/db_xref="SPTREMBL:P96868"
/translation="MQPSHPTRPGAVIRYVGSSLDTCPMTTFAGKTAASADKVRGGYY
TPPAVARFLAHWVHQAGPKILEPSCGDGRILRELSAITDHAHGVELVAREAKKSRDFA
SVDTENLFTWLHKTQLGSWDGVAGNPPYIRFGNWASEQRDPALELMRRVGLRPTKLTN
AWVPFVVASTTLARDGGRVGLVVPAELLQVTYAAQLREFLLSRYREITLVTFERLVFD
GILQEVVLFCGVVGPGPAHIRTVRLGDANDLNALGDKDFTNESAPALLHEKEKWTKYF
LDPAQIRLLRGLKQSATMIRLGELADVDVGIVTGRNSFFTFTDAKAQALGLRAHCVPL
VSRSAQLSGLIYDEDCRACDVAGNHRTWLLDAADYPTDPALVAHITAGEAAGVHLGYK
CSIRKPWWSTPSLWMPDLFMLRQTHFAPRLTVNAAAATSTDTVHRVRLDPNVDPATLA
AVFHNSATFAFAEIMGRSYGGGILELEPREAEQLPMPPPAYGSAELAQDVDLLLKANE
IDKALDVVDRHVLIDGLGLSPRLVAGCRAAWLTLRDRRTKRGSRR"
misc_feature 3225. .3245
/gene="Rv3263"
/note="PS00092 N-6 Adenine-specific DNA methylases signature"
gene complement(4577. .5656)
/gene="rmlA2"
CDS complement(4577. .5656)
/gene="rmlA2"
/note="Rv3264c, (MTCY71.04c), len: 359, rfbF transferase, similar to MPG1_YEAST P41940 mannose-1-phosphate guanyltransferase (361 aa), fasta scores, opt: 557, E(): 1.9e-27, (28.9% identity in 353 aa overlap), also similar to RBA1_ECOLIP37744, glucose-1-phosphate thymidylyltransferase (293 aa),fasta scores, opt: 317, E(): 7.7e-13, (26.4% identity in 235 aa overlap)"
/codon_start=1
/transl_table=11
/product="rmlA2"
/protein_id="CAB07091.1"
/db_xref="GI:1877319"
/db_xref="SPTREMBL:P96869"
/translation="MATHQVDAVVLVGGKGTRLRPLTLSAPKPMLPTAGLPFLTHLLS
RIAAAGIEHVILGTSYKPAVFEAEFGDGSALGLQIEYVTEEHPLGTGGGIANVAGKLR
NDTAMVFNGDVLSGADLAQLLDFHRSNRADVTLQLVRVGDPRAFGCVPTDEEDRVVAF
LEKTEDPPTDQINAGCYVFERNVIDRIPQGREVSVEREVFPALLADGDCKIYGYVDAS
YWRDMGTPEDFVRGSADLVRGIAPSPALRGHRGEQLVHDGAAVSPGALLIGGTVVGRG
AEIGPGTRLDGAVIFDGVRVEAGCVIERSIIGFGARIGPRALIRDGVIGDGADIGARC
ELLSGARVWPGVFLPDGGIRYSSDV"
gene complement(5658. .6563)
/gene="wbbL"
CDS complement(5658. .6563)
/gene="wbbL"
/note="Rv3265c, (MTCY71.05c), len: 301, unknown but similar to YEFJ_ECOLI P36667 hypothetical 31.1 kd protein in gnd-r (264 aa), fasta scores, opt: 253, E(): 9.8e-10, (28.6% identity in 259 aa overlap)"
/codon_start=1
/transl_table=11
/product="wbbL"
/protein_id="CAB07092.1"
/db_xref="GI:1877320"
/db_xref="SPTREMBL:P96870"
/translation="MVAVTYSPGPHLERFLASLSLATERPVSVLLADNGSTDGTPQAA
VQRYPNVRLLPTGANLGYGTAVNRTIAQLGEMAGDAGEPWVDDWVIVANPDVQWGPGS
IDALLDAASRWPRAGALGPLIRDPDGSVYPSARQMPSLIRGGMHAVLGPFWPRNPWTT
AYRQERLEPSERPVGWLSGSCLLVRRSAFGQVGGFDERYFMYMEDVDLGDRLGKAGWL
SVYVPSAEVLHHKAHSTGRDPASHLAAHHKSTYIFLADRHSGWWRAPLRWTLRGSLAL
RSHLMVRSSLRRSRRRKLKLVEGRH"
gene complement(6574. .7488)
/gene="rmlD"
CDS complement(6574. .7488)
/gene="rmlD"
/note="Rv3266c, (MTCY71.06c), len: 304, rfbD, similar to eg STRL_STRGR P29781 dtdp-4-dehydrorhamnose reductase (304 aa), fasta scores, opt: 788, E(): 0, (47.4% identity in 304 aa overlap)"
/codon_start=1
/transl_table=11
/product="rmlD"
/protein_id="CAB07093.1"
/db_xref="GI:1877321"
/db_xref="SPTREMBL:P96871"
/translation="MAGRSERLVITGAGGQLGSHLTAQAAREGRDMLALTSSQWDITD
PAAAERIIRHGDVVINCAAYTDVDGAESNEAVAYAVNATGPQHLARACARVGARLIHV
STDYVFDGDFGGAEPRPYEPTDETAPQGVYARSKLAGEQAVLAAFPEAAVVRTAWVYT
GGTGKDFVAVMRRLAAGHGRVDVVDDQTGSPTYVADLAEALLALADAGVRGRVLHAAN
EGVVSRFGQARAVFEECGADPQRVRPVSSAQFPRPAPRSSYSALSSRQWALAGLTPLR
HWRSALATALAAPANSTSIDRRLPSTRD"
gene 7564. .9060
/gene="Rv3267"

Query Match 72.8%; Score 18.2; DB 3; Length 42729;

Best Local Similarity 87.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||| ||||||||||||| | |
Db    8563 GATCCGCGTCAGGTGCAGGACTG 8585
```

Search completed: October 9, 2001, 12:14:40
Job time: 3995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:01 ; Search time 470.56 Seconds
(without alignments)
33.359 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	20	AAX01303	E. coli biotin syn
2	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
3	25	100.0	1121	7	AAN60496	Sequence encoding
4	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
5	18.2	72.8	906	11	AAQ04701	Sequence encoding
6	18.2	72.8	979	12	AAQ12149	Cytotoxic lymphocy
7	18.2	72.8	984	18	AAT50759	Sequence of ovine
8	18.2	72.8	987	16	AAQ97484	Human interleukin-
9	18.2	72.8	987	19	AAV07273	Human Interleukin-
10	18.2	72.8	987	19	AAV07217	DNA encoding human
11	18.2	72.8	987	20	AAZ40404	Human interleukin
12	18.2	72.8	987	20	AAV83199	Interleukin (IL)-1
13	18.2	72.8	1018	15	AAQ67879	EPV 42 kD promoter
14	18.2	72.8	1018	20	AAZ08510	EPV 42 kDa/human I
15	18.2	72.8	1019	21	AAA70709	Bovine interleukin
16	18.2	72.8	1389	22	AAF76880	Human p40-IL-2 cod
17	18.2	72.8	1397	18	AAV00401	40 kDa subunit of
18	18.2	72.8	1397	18	AAT77849	40 kDa subunit of
19	18.2	72.8	1397	18	AAT97969	40 kDa subunit of
20	18.2	72.8	1397	19	AAV37767	Cytotoxic lymphocy
21	18.2	72.8	1399	21	AAZ40021	Human p40 coding s
c 22	18.2	72.8	2096	21	AAZ50892	Human receptor-ass
23	18.2	72.8	2300	21	AAA92406	Human interleukin
24	18.2	72.8	2362	13	AAQ23586	Natural killer cel
25	18.2	72.8	2362	17	AAT48098	Human interleukin-
26	18.2	72.8	2362	19	AAV42535	Human interleukin-
27	18.2	72.8	2362	19	AAV15227	Heterodimeric huma
28	18.2	72.8	2362	21	AAZ28862	Human interleukin-
29	18.2	72.8	2362	22	AAF56280	IL-12 vaccine adju
30	18.2	72.8	5686	20	AAZ40415	Plasmid pIN1143 co
31	18.2	72.8	6139	21	AAZ40023	Interleukin-12 fus
c 32	17.2	68.8	586	21	AAC75563	Human ORFX ORF1118
c 33	16.8	67.2	5269	21	AAC76615	Human ORFX ORF2170
34	16.6	66.4	34	16	AAQ75832	Sense primer to am
35	16.6	66.4	220	22	AAF28005	Human TR13 coding
36	16.6	66.4	220	22	AAF28044	Human TR13 coding
37	16.6	66.4	625	21	AAC98199	Human colon cancer
38	16.6	66.4	651	21	AAF14300	Aspergillus oryzae
39	16.6	66.4	681	22	AAF71312	Corynebacterium gl
40	16.6	66.4	924	20	AAX18177	Canine mature CLAF
41	16.6	66.4	990	19	AAV13816	Canine IL-12 P40 s
42	16.6	66.4	990	19	AAV13817	Canine IL-12 P40 s
43	16.6	66.4	990	20	AAX35615	Nucleic acid encod
44	16.6	66.4	990	20	AAX35625	Nucleic acid encod
45	16.6	66.4	990	20	AAX03162	Nucleic acid encod

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.024;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 cgatcctcgtcaggtgcaggtcagc 129


RESULT   2
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    128 cgatcctcgtcaggtgcaggtcagc 152


RESULT   3
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 0.025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    146 cgatcctcgtcaggtgcaggtcagc 170


RESULT   4
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
```

```
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.029;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 cgatcctcgtcaggtgcaggtcagc 245


RESULT   5
AAQ04701
ID   AAQ04701 standard; DNA; 906 BP.
XX
AC   AAQ04701;
XX
DT   11-OCT-1990  (first entry)
XX
DE   Sequence encoding natural killer cell stimulatory factor (NKSF).
XX
KW   Natural killer cell stimulatory factor; NKSF; cancer;
KW   leukopaenia; anaemia; radiation; bone marrow transplantation; ds.
XX
OS   Homo sapiens.
XX
```

```
PN   WO9005147-A.
XX
PD   17-MAY-1990.
XX
PF   09-NOV-1989;   89WO-0005027.
XX
PR   10-NOV-1988;   88US-0269945.
XX
PA   (GENE-) GENETICS INST INC.
PA   (WIST-) WISTAR INST.
XX
PI   Trinchieri G,  Perussia B,  Kobayashi M,  Clark SC,  Wong GG;
XX
DR   WPI; 1990-178800/23.
DR   P-PSDB; AAR05171.
XX
PT   New natural killer cell stimulatory factor -
PT   useful for treating cancer and eg. leukopenia, bacterial and
PT   viral infections, anaemia and B- or T-cell deficiencies.
XX
PS   Claim 16; Page 60; 69pp; English.
XX
CC   Natural killer cell stimulatory factor (NKSF) is useful in
CC   treating cancer and other disease states which respond to NK cell
CC   activity. Such states may arise from immune or haematopoietic-cell
CC   deficiency following a bone marrow translpantation, exposure to
CC   radiation and disease.
XX
SQ   Sequence 906 BP; 244 A; 220 C; 238 G; 204 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 906;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT   6
AAQ12149
ID   AAQ12149 standard; DNA; 979 BP.
XX
AC   AAQ12149;
XX
DT   10-SEP-1991  (first entry)
XX
DE   Cytotoxic lymphocyte maturation factor (40 kD subunit) gene.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; probe; antitumour;
KW   monoclonal antibodies; T cell; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..979
FT                   /*tag=  a
FT                   /product= CLMF
FT   sig_peptide     41..106
FT                   /*tag=  b
FT                   /note= "hydrophobic signal peptide"
FT   mat_peptide     107..979
FT                   /*tag=  c
XX
PN   EP433827-A.
XX
PD   26-JUN-1991.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN-LA ROCHE AG.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD,  Pan YCE;
PI   Podlaski FJ,  Stern AS;
XX
DR   WPI; 1991-186715/26.
DR   P-PSDB; AAR12493.
XX
PT   New cytotoxic lymphocyte maturation factor - used as antitumour
PT   agent and monoclonal antibodies used in transplantation to block
PT   proliferation and activation of cytotoxic T cells
XX
PS   Disclosure; Fig 25 (a-c); 90pp; English.
XX
CC   The natural CLMF protein is a 75 kD heterodimer comprising two
CC   polypeptide subunits, a 40 kD and a 35 kD subunit (AAQ12150).
CC   The subunits are bonded together via one or more disulfide bonds.
CC   The CLMF can interact with IL-2 to synergistically induce the
CC   cytolytic activity of Lymphokine Activated Killer (LAK) cells.  It is
CC   also capable of stimulating T-cell growth.  It can be used for the
CC   prepn. of LAK- and T-cell activator and antitumour compsns. and for
CC   stimulating Natural Killer (NK) cells.
XX
SQ   Sequence 979 BP; 261 A; 245 C; 257 G; 216 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 12;  Length 979;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT   7
AAT50759
ID   AAT50759 standard; cDNA; 984 BP.
XX
AC   AAT50759;
XX
DT   24-SEP-1997  (first entry)
XX
DE   Sequence of ovine IL-12 40 kD sununit cDNA.
XX
KW   Cytokine; ovine; sheep; interleukin-5; interleukin-12; IL-5; IL-12;
KW   livestock; cow; stress; transport; vaccine adjuvant; veterinary;
KW   cancer; immunosuppression; allergy; reproductive system; growth;
KW   early maturity; antibody; diagnosis; immunopotentiator;
KW   early haematopoietic progenitor cell; cytotoxic cell; thymocyte;
KW   secretion; IgM; IgA; bacterial endotoxin; gamma-interferon; ss.
XX
OS   Ovis ammon aries.
XX
PN   WO9700321-A1.
XX
PD   03-JAN-1997.
XX
PF   14-JUN-1996;   96WO-AU00360.
XX
PR   27-OCT-1995;   95AU-0006244.
PR   14-JUN-1995;   95AU-0003502.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Seow H,  Wood P;
XX
DR   WPI; 1997-077528/07.
DR   P-PSDB; AAW08483.
XX
PT   Nucleic acid encoding ovine interleukin-5 or -12 - used as vaccine
```

PT adjuvants and to treat or prevent microbial infections in livestock
XX
PS Claim 10; Page 49-51; 78pp; English.
XX
CC The sequences given in AAT50757-59 encode ovine interleukin-12 (IL-12)
CC 35 kD subunit (partial and full length sequence) and the 40 kD subunit.
CC Ovine IL-5 or IL-12 are used to treat and/or prevent infections in
CC livestock (esp. cows and sheep), particularly where the animals are
CC stressed, e.g. during transport. IL-5 and IL-12 can also be used as
CC adjuvants in vaccines for veterinary use (partic. weakly immunogenic
CC subunit or synthetic peptide vaccines). They may also be used to
CC treat cancer, immunosuppression and allergy, to enhance/suppress the
CC reproductive system and to promote growth or early maturity. Optionally
CC interleukin can be delivered from constructs or delivery cells and
CC antibodies are useful in enzyme immunoassays for rapid diagnosis of
CC infection. The interleukins are immunopotentiators, especially IL-5
CC promotes growth of early haematopoietic progenitor cells and generation
CC of cytotoxic cells from thymocytes, also it stimulates production and
CC secretion of IgM and IgA (in synergism with bacterial endotoxin).
CC IL-12 induces production of gamma-interferon by, and proliferation
CC of, T and NK cells and increases the (non-)specific cytolytic
CC lymphocyte response. The genetic constructs can also be used for
CC in vitro production of IL-5 or -12.
XX
SQ Sequence 984 BP; 273 A; 243 C; 253 G; 215 T; 0 other;

Query Match 72.8%; Score 18.2; DB 18; Length 984;
Best Local Similarity 87.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 746 attctcggcaggtggaggtcagc 768

RESULT 8
AAQ97484
ID AAQ97484 standard; cDNA; 987 BP.
XX
AC AAQ97484;
XX
DT 22-DEC-1995 (first entry)
XX
DE Human interleukin-12 40 kD subunit.
XX
KW Interleukin-12; IL-12; cytokine; gamma-interferon; T cell;
KW natural killer cell; graft vs. host disease; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..987
FT /*tag= a
XX
PN WO9519786-A.
XX
PD 27-JUL-1995.
XX
PF 20-JAN-1995; 95WO-US00879.
XX
PR 21-JAN-1994; 94US-0186529.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (GEMY) GENETICS INST INC.
XX
PI Sykes M, Wolf SF;
XX
DR WPI; 1995-269273/35.
DR P-PSDB; AAR79186.
XX
PT Control of graft versus host disease using interleukin 12 - partic.
PT admin. at the time of bone marrow transplant
XX
PS Disclosure; Page 21-22; 32pp; English.
XX
CC AAQ97484 and AAQ97485 are from human lymphoblast cell line RMPI 8866.
CC IL- 12 is a heterodimeric cytokine which was originally identified as a
CC factor which induces gamma-interferon from T cells and natural
CC killer cells as set forth in PCT/US91/06332, published April 2 1992,
CC and incorporated by reference. EP 433827, published June 26 1991,
CC disclosed IL-12 as a cytotoxin lymphocyte maturation factor. In the
CC patent, IL-12 is defined as a heterodimeric glycoprotein comprised
CC of two covalently linked subunits having mol. wts of about 40 kD and
CC 35 kD. The two subunits may be disulfide-bonded. IL-12 is
CC administered to prevent or ameliorate GVHD. Dosage is 1 ng-0.1 mg
CC (esp. 1.0 ng-1 microg)/kg daily for 3 days, the first dose given
CC 1 hr. after transplant. IL-12 can be used as the native heterodimer
CC or as subunits.
XX
SQ Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;

Query Match 72.8%; Score 18.2; DB 16; Length 987;
Best Local Similarity 87.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 743 attctcggcaggtggaggtcagc 765

RESULT 9
AAV07273
ID AAV07273 standard; cDNA; 987 BP.
XX
AC AAV07273;
XX
DT 25-SEP-1998 (first entry)
XX
DE Human Interleukin-12 p40 gene.
XX
KW IL-12 subunit; expression construct; treatment; asthma; microbial
KW infection; viral infection; cancer; Human; Interleukin; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..987
FT /*tag= a
FT /product= "IL-12 p40"
XX
PN WO9817689-A2.
XX
PD 30-APR-1998.
XX
PF 10-OCT-1997; 97WO-US18779.
XX
PR 18-OCT-1996; 96US-0028676.
XX
PA (GENE-) GENEMEDICINE INC.
XX
PI Deshpande D, Freimark B, Nordstrom J;
XX
DR WPI; 1998-261428/23.
DR P-PSDB; AAW51311.
XX
PT Constructs for expression of interleukin-12 sub-units - are used
PT for delivery of IL-12 sub-units for treating e.g. asthma, microbial
PT or viral infections and certain cancers
XX
PS Claim 1; Page 34; 80pp; English.
XX
CC The IL-12 subunits AAV07273 and AAV07274 can be expressed using an
CC appropriate expression construct engineered into a plasmid. The
CC plasmids can provide for efficient expression of IL-12 subunits. The

```
CC   products can be used for the treatment of asthma, microbial and viral
CC   infections and certain cancers.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  10
AAV07217
ID   AAV07217 standard; DNA; 987 BP.
XX
AC   AAV07217;
XX
DT   12-AUG-1998  (first entry)
XX
DE   DNA encoding human interleukin-12 40kDa subunit.
XX
KW   Interleukin-12; IL-12; graft-versus-host disease; GVHD.
KW   bone marrow transplantation; graft-versus-leukemia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..987
FT                   /*tag=  a
FT                   /product= Interleukin-12_40kDa_subunit
XX
PN   US5756085-A.
XX
PD   26-MAY-1998.
XX
PF   30-APR-1996;   96US-0640386.
XX
PR   30-APR-1996;   96US-0640386.
PR   21-JAN-1994;   94US-0186529.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sykes M,  Wolf SF;
XX
DR   WPI; 1998-321459/28.
DR   P-PSDB; AAW51221.
XX
PT   Therapy of graft-versus-host disease  - by administration of
PT   interleukin-12 to maintain and/or promote graft-versus-leukaemia
PT   effects
XX
PS   Disclosure; Columns 17-20; 20pp; English.
XX
CC   The invention relates to a method for the treatment or prevention
CC   of graft-versus-host disease (GVHD) which comprises the administration
CC   of interleukin-12 (IL-12) at the time of bone-marrow transplantation to
CC   maintain or promote graft-versus-leukaemia effects. Preferably, dosage is
CC   0.001-100 (preferably 1-100) mu g/kg for 3 days starting on the day of
CC   the bone marrow transplant. IL-12 is administered daily until there is
CC   an improvement in the condition. Administration is oral, cutaneous,
CC   subcutaneous, intravenous or by inhalation preferably subcutaneous, or
CC   intravenous. The present sequence encodes the human interleukin-12
CC   40kDa subunit.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  11
AAZ40404
ID   AAZ40404 standard; DNA; 987 BP.
XX
AC   AAZ40404;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Human interleukin 12 p40 subunit wild type coding sequence.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
DR   P-PSDB; AAY42487.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Claim 11; Page 43-45; 137pp; English.
XX
CC   This sequence represents the wild type coding region for the human
CC   interleukin 12 p40 subunit. The invention relates to a novel plasmid
CC   comprising a cytomegalovirus (CMV) promoter transcriptionally linked
CC   with an interferon alpha (IFN-alpha) coding sequence, and a growth
CC   hormone 3'-untranslated region (UTR).  The plasmids can be used for
CC   treating a mammalian condition or disease, particularly cancer.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 20;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  12
AAV83199
ID   AAV83199 standard; DNA; 987 BP.
XX
AC   AAV83199;
XX
DT   02-MAR-1999  (first entry)
XX
DE   Interleukin (IL)-12B cDNA.
```

```
XX
KW   Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW   immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KW   Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
KW   human immunodeficiency virus; ds.
XX
OS   Homo sapiens.
XX
PN   WO9844788-A2.
XX
PD   15-OCT-1998.
XX
PF   09-APR-1998;   98WO-US06944.
XX
PR   09-DEC-1997;   97US-0069163.
PR   09-APR-1997;   97US-0838702.
PR   01-MAY-1997;   97US-0848760.
XX
PA   (CHAN/) CHANG L.
XX
PI   Chang L;
XX
DR   WPI; 1999-024005/02.
XX
PT   Use of immunodeficient mice comprising human cells - particularly
PT   SCTD/beige mice comprising human immune cells for evaluating
PT   vaccines against cancers or human pathogens, e.g. HIV
XX
PS   Example 10(c)(i); Page 110; 154pp; English.
XX
CC   Immunodeficient mice comprising human cells can be used for
CC   exposure to human pathogens and/or their components or human
CC   tumour cells and human peripheral blood lymphocytes. Also claimed
CC   is a vaccine comprising a cell modified to express an antigen and
CC   an immune-modulating protein, this is preferably an expression
CC   vector comprising a polynucleotide sequence that encodes the
CC   antigen and immune-modulating protein.  Such vectors can be used
CC   to treat a subject having a tumour by transferring the expression
CC   vector into the tumour so that the antigen and the immune-modulator
CC   are expressed by at least the tumour. The methods can be used for
CC   producing and evaluating vaccines including cancer vaccines and
CC   vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC   Mycobacterium, Listeria or Plasmodium.  Two primers (AAV83197,
CC   AAV83198) were used to amplify the IL-12B for insertion into
CC   expression vectors of the invention.
XX
SQ   Sequence 987 BP; 264 A; 245 C; 257 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 20;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  13
AAQ67879
ID   AAQ67879 standard; DNA; 1018 BP.
XX
AC   AAQ67879;
XX
DT   23-MAR-1995  (first entry)
XX
DE   EPV 42 kD promoter/Human IL-12 p40 subunit expression cassette.
XX
KW   Polymerase chain reaction; primer; amplify; NYVAC; ALVAC; recombinant;
KW   murine; interleukin-2; IL-2; pRW825; pmut-1; pBS-SK; pMM151; TK vector;
KW   plasmid; vaccinia; H6 promoter; amplify; primer; antigenic response;
KW   polymerase chain reaction; poxvirus; pSD542; immunological response;
KW   pathogen; human; interferon; IFN; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..31
FT                   /*tag=  a
FT                   /note= "EPV 42 kD promoter"
FT   CDS             32..1018
FT                   /*tag=  b
FT                   /product= Human IL-12 p40
XX
PN   WO9416716-A.
XX
PD   04-AUG-1994.
XX
PF   21-JAN-1994;   94WO-US00888.
XX
PR   21-JAN-1993;   93US-0007115.
PR   19-JAN-1994;   94US-0184009.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Cox WI,  Paoletti E,  Tartaglia J;
XX
DR   WPI; 1994-263767/32.
XX
PT   Attenuated recombinant virus used for cancer therapy - comprises
PT   DNA encoding cytokine and/or tumour associated antigen
XX
PS   Example 27; Fig 33; 232pp; English.
XX
CC   The sequences given in AAQ67879-80 are expression cassettes which
CC   contain the p40 and p35 subunits of human interleukin-12 (IL-12) under
CC   the control of the entomopox virus (EPV) promoter and the vaccinia E3L
CC   promoter, respectively.  These sequences were used in the construction
CC   of an ALVAC-based recombinant virus containing the genes for the two
CC   subunits of human IL-12.  The p40 subunit was amplified using primers
CC   JP202 and JP189 and placed under the control of the entomopox virus
CC   (EPV) 42 kD promoter in plasmid PBSHIL12p40II.  The p35 subunit was
CC   amplified using primers JP186 and JP201 and placed under the control
CC   of the vaccinia E3L promoter using primers JP218 and JP220.  The
CC   expression cassette was isolated and ligated with pBSSK+ generating
CC   plasmid PBSHIL12p35II.  The p35 and p40 expression cassettes were
CC   excised and assembled in pBSSK+ to give PBSHIL12 in which the two
CC   expression cassettes are orientated head-to-head relative to each other.
CC   The combination casette containing the poxvirus promoted genes for both
CC   IL-12 subunits was excised from PBSHIL12 and cloned in to ALVAC C6
CC   vector plasmid pC6L to give pC6HIL12.  Recombination between this donor
CC   plasmid and ALVAC rescuing virus produced recombinant virus ALVAC+IL12,
CC   which contains both the human IL-12 genes in the C6 locus.  The resulting
CC   virus may be used in a composition for inducing an antigenic or
CC   immunological response, ie. for immunisation against pathogens.
XX
SQ   Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 15;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 attctcggcaggtggaggtcagc 796


RESULT  14
AAZ08510
ID   AAZ08510 standard; DNA; 1018 BP.
XX
AC   AAZ08510;
XX
DT   19-OCT-1999  (first entry)
XX
```

DE EPV 42 kDa/human IL-12 P40 expression cassette.
XX
KW Attenuated recombinant virus; cytokine; tumour associated antigen;
KW NYVAC recombinant virus; ALVAC recombinant virus; gene therapy; rabies;
KW cancer; tumour necrosis factor; nuclear phosphoprotein; p53; IL-2; GMCSF;
KW interleukin; interferon; IFN-gamma; IL-4; melanoma associated antigen;
KW carcinoembryonic antigen; immunisation; antigenic; poxvirus; influenza;
KW immunological response; immunotherapy; vaccine; Newcastle Disease; ss.
XX
OS Synthetic.
OS Entomopox virus.
OS Homo sapiens.
XX
PN US5942235-A.
XX
PD 24-AUG-1999.
XX
PF 02-JUN-1995; 95US-0458356.
XX
PR 02-JUN-1995; 95US-0458356.
PR 24-DEC-1981; 81US-0334456.
PR 08-DEC-1982; 82US-0446824.
PR 19-JUN-1984; 84US-0622135.
PR 27-AUG-1987; 87US-0090209.
PR 28-AUG-1987; 87US-0090711.
PR 20-OCT-1987; 87US-0110335.
PR 25-APR-1988; 88US-0186054.
PR 23-AUG-1988; 88US-0234390.
PR 08-MAR-1989; 89US-0320471.
PR 14-FEB-1990; 90US-0478179.
PR 14-JUN-1990; 90US-0537882.
PR 14-JUN-1990; 90US-0537890.
PR 07-JAN-1991; 91US-0638080.
PR 07-MAR-1991; 91US-0666056.
PR 11-JUN-1991; 91US-0713967.
PR 16-DEC-1991; 91US-0805567.
PR 03-MAR-1992; 92US-0847977.
PR 06-MAR-1992; 92US-0847951.
PR 04-MAY-1992; 92US-0881995.
PR 22-JUL-1992; 92US-0918278.
PR 20-JAN-1993; 93US-0007115.
PR 19-JAN-1994; 94US-0184009.
PR 14-APR-1994; 94US-0228926.
PR 13-SEP-1994; 94US-0306259.
XX
PA (HEAL-) HEALTH RES INC.
XX
PI Paoletti E;
XX
DR WPI; 1999-493494/41.
XX
PT Recombinant poxviruses comprising exogenous DNA encoding antigenic
PT determinants useful in immunotherapy to immunize against cancers and
PT other diseases such as influenza, Newcastle Disease and rabies
XX
PS Example 27; Fig 33; 163pp; English.
XX
CC The present invention describes a recombinant poxvirus (I), comprising
CC exogenous DNA encoding an antigenic determinant of a pathogen which is
CC then expressed in vivo in infected host cells after administration to a
CC patient and therefore induces an immunological response. (I) may be used
CC to vaccinate patients against a wide range of diseases and disorders
CC depending on the type of antigen encoded by the exogenous DNA. (I) may
CC be used to vaccinate against diseases such as rabies, influenza and
CC Newcastle Disease. It is particularly useful for immunising against
CC cancers. The poxvirus (I) also provides a means of manipulating
CC lymphocytes and tumour cells for use in cell-based immunotherapeutic
CC modalities for cancer. (I) also have enhanced safety compared to
CC unattenuated viruses (attenuation reduces the virulence of the viruses)
CC and known recombinant poxvirus vaccines. This increased level of safety
CC reduces the possibility of a 'runaway' infection in the host and reduces
CC the chance of transmission from vaccinated to unvaccinated individuals
CC and contamination of the environment. The present sequence represents a
CC EPV 42 kDa/human IL-12 P40 expression cassette used in the
CC exemplification of the present invention.
XX
SQ Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;

```
  Query Match             72.8%;  Score 18.2;  DB 20;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 attctcggcaggtggaggtcagc 796
```

RESULT 15
AAA70709
ID AAA70709 standard; DNA; 1019 BP.
XX
AC AAA70709;
XX
DT 15-DEC-2000 (first entry)
XX
DE Bovine interleukin-12 p40 subunit gene.
XX
KW Antiinfective; bovine; interleukin 12; IL12; heterodimer; infection; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 12..995
FT /*tag= a
FT /product= "IL-12 p40 subunit"
XX
PN JP2000157274-A.
XX
PD 13-JUN-2000.
XX
PF 20-NOV-1998; 98JP-0331052.
XX
PR 20-NOV-1998; 98JP-0331052.
XX
PA (DAUC) DAIICHI PHARM CO LTD.
XX
DR WPI; 2000-468201/41.
DR P-PSDB; AAB15391.
XX
PT Preparation of bovine interleukin 12 comprising using a vector
PT containing the gene encoding it, useful for the prevention and
PT treatment of chronic and opportunistic infections -
XX
PS Example 1; Page 8-9; 11pp; Japanese.
XX
CC The invention relates to a method for the preparation of bovine
CC interleukin 12 (BoIL12). BoIL12 consists of a heterodimer of p35
CC and p40 subunits. BoIL12 is produced by introducing into a host cell,
CC recombinant vectors containing genes encoding the p35 and p40 subunits.
CC This sequence represents the coding region for the bovine IL12 p40
CC subunit. Interleukin 12 is used for the prevention and treatment of
CC chronic infections and opportunistic infections.
XX
SQ Sequence 1019 BP; 273 A; 248 C; 271 G; 227 T; 0 other;

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 1019;
  Best Local Similarity   87.0%;  Pred. No. 28;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     757 attctcggcaggtggaggtcagc 779
```

```
RESULT  16
AAF76880
ID   AAF76880 standard; DNA; 1389 BP.
XX
AC   AAF76880;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human p40-IL-2 coding sequence.
XX
KW   Human; p40-IL-2; interleukin-2; IL-2; antiviral; antibacterial;
KW   antifungal; cytostatic; antiparasitic; vaccine; gene therapy;
KW   fusion protein; protein complex; cytokine; immunoglobulin; cancer;
KW   infection; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200110912-A1.
XX
PD   15-FEB-2001.
XX
PF   09-AUG-2000; 2000WO-US21715.
XX
PR   09-AUG-1999;   99US-0147924.
XX
PA   (LEXI-) LEXIGEN PHARM CORP.
XX
PI   Gillies SD,  Lo KM;
XX
DR   WPI; 2001-191523/19.
XX
PT   Novel multifunctional fusion protein or protein complexes useful for
PT   treating cancer and viral infections, comprise two different cytokine
PT   molecules and a targeting group -
XX
PS   Example 1; Page 75-76; 90pp; English.
XX
CC   The present sequence is given in a specification relating to
CC   multifunctional fusion proteins and protein complexes comprising at
CC   least two different cytokine molecules and a targeting group such as an
CC   immunoglobulin. The fusion proteins and protein complexes are useful for
CC   treating cancers, and bacterial, viral, fungal and parasitic infections.
CC   They are also useful in gene therapy and as vaccines to prevent or treat
CC   various diseases.
XX
SQ   Sequence 1389 BP; 413 A; 325 C; 327 G; 324 T; 0 other;

  Query Match             72.8%;  Score 18.2;  DB 22;  Length 1389;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  17
AAV00401
ID   AAV00401 standard; DNA; 1397 BP.
XX
AC   AAV00401;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790309-A1.
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes J,  Pan YE;
PI   Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404734/38.
DR   P-PSDB; AAW36941.
XX
PT   Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
PT   for producing CLMF protein and as an antitumour agent
XX
PS   Disclosure; Fig 25; 83pp; English.
XX
CC   The present DNA sequence encodes a novel 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAV00402 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. The subunits may be used for preparing
CC   the CLMF protein and for preparing medicants for antitumour therapy.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;

  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT  18
AAT77849
ID   AAT77849 standard; DNA; 1397 BP.
XX
AC   AAT77849;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790255-A2.
```

```
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD;
PI   Pan YE,  Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404698/38.
DR   P-PSDB; AAW24235.
XX
PT   Antibodies to cytotoxic lymphocyte maturation factor protein -
PT   useful as affinity ligands, assay reagents and medicaments
XX
PS   Disclosure; Fig 25; 80pp; English.
XX
CC   The present DNA sequence encodes the 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAT77850 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. A novel antibody directed against an epitope
CC   of CLMF protein (see AAW24237 for epitope) is capable of neutralising
CC   and/or inhibiting CLMF bioactivity. This antibody may be used as an
CC   affinity ligand for purifying the CLMF protein. The antibody can also
CC   be used as an assay reagent for detecting the CLMF protein, and as a
CC   medicament for selectively blocking the proliferation and activation of
CC   cytotoxic T cells, e.g. in transplantation.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT  19
AAT97969
ID   AAT97969 standard; DNA; 1397 BP.
XX
AC   AAT97969;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790308-A1.
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD;
PI   Pan YE,  Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404733/38.
DR   P-PSDB; AAW36939.
XX
PT   Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
PT   for producing protein and as antitumour agent
XX
PS   Disclosure; Fig 25; 83pp; English.
XX
CC   The present DNA sequence encodes a 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAT97970 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. The subunits may be used for preparing
CC   the CLMF protein and for preparing medicants for antitumour therapy.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT  20
AAV37767
ID   AAV37767 standard; cDNA; 1397 BP.
XX
AC   AAV37767;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Cytotoxic lymphocyte maturation factor 40 kD subunit encoding cDNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; cytokine; synergize;
KW   interleukin 2; IL-2; human; lymphoblastoid cell; detection; purifying;
KW   proliferation; cytotoxic T cell; transplantation; antibody; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
FT                   /product= "CLMF 40 kD subunit"
XX
PN   US5780597-A.
XX
PD   14-JUL-1998.
XX
PF   02-JUN-1995;   95US-0460061.
XX
PR   24-MAR-1992;   92US-0857023.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
PR   27-AUG-1990;   90US-0572284.
```

```
PR   02-MAR-1994;   94US-0205011.
PR   02-JUN-1995;   95US-0460061.
XX
PA   (HOFF ) HOFFMANN LA ROCHE INC.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD;
PI   Pan YE,  Podlaski FJ,  Stern AS;
XX
DR   WPI; 1998-413150/35.
DR   P-PSDB; AAW58725.
XX
PT   New antibodies to cytotoxic lymphocyte maturation factor - useful
PT   for detecting, purifying, and/or blocking proliferation and
PT   activation of cytotoxic T cells, such as in transplantation(s)
XX
PS   Example 10; Fig 25; 71pp; English.
XX
CC   An isolated antibody has been developed which binds specifically to
CC   cytotoxic lymphocyte maturation factor (CLMF) and neutralises CLMF
CC   induced proliferation by more than 50% in a CLMF dependent T cell
CC   growth assay and/or inhibits binding of at least 60% of the factor
CC   to phytohaemagglutinin (PHA) activated peripheral blood lymphocyte
CC   (PBL) blasts as determined in a CLMF receptor binding assay. CLMF is
CC   a heterodimeric protein having a molecular weight band of 75 kD,
CC   determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis
CC   (SDS-PAGE) under non-reducing and/or reducing conditions, providing a
CC   first subunit having a molecular weight band of 40 kD and a second
CC   subunit having a molecular weight of 35 kD. The present sequence
CC   encodes the CLMF 40 kD subunit, from an example of the present
CC   invention. The antibody can be used for the purification and/or
CC   detection of CLMF. It is also used in therapeutic treatments which
CC   require selective blocking of proliferation and activation of
CC   cytotoxic T cells (CTLs) such as in transplantation.
XX
SQ   Sequence 1397 BP; 390 A; 309 C; 333 G; 365 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 19;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805
```

```
RESULT  21
AAZ40021
ID   AAZ40021 standard; DNA; 1399 BP.
XX
AC   AAZ40021;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Human p40 coding sequence.
XX
KW   Interleukin-12; IL-12; fusion protein; IL-12 p35 subunit; B7 protein;
KW   IL-12 p40 subunit; gene therapy; tumour; leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   US5994104-A.
XX
PD   30-NOV-1999.
XX
PF   08-NOV-1996;   96US-0751767.
XX
PR   08-NOV-1996;   96US-0751767.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Anderson RJ,  Prentice HG,  MacDonald ID;
XX
DR   WPI; 2000-038261/03.
DR   P-PSDB; AAY54919.
XX
PT   Nucleic acid constructs encoding interleukin-12 fusion proteins useful
PT   for treating leukemia and other cancers -
XX
PS   Example; Fig 9; 73pp; English.
XX
CC   This sequence encodes the human p40 protein sequence.
CC   The invention relates to an isolated nucleic acid construct (I)
CC   comprising a region encoding an interleukin-12 (IL-12) fusion protein
CC   (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and a linker
CC   peptide (joining the subunits)) and a region encoding a B7 protein. (I)
CC   may be used to produce IL-12 fusion proteins according to standard
CC   recombinant DNA methodologies. The fusion proteins may be produced either
CC   in vitro in a fermentation culture or in vivo as part of a gene therapy
CC   protocol (in this case (I) is used to transform a patients cells, which
CC   then secrete the functional polypeptide to supplement the patients own
CC   production of IL-12 or to rectify mutations which lead to the expression
CC   of inactive polypeptides). The fusion proteins produced in this way may
CC   be used to treat any disease which responds to IL-12 such as tumours
CC   (both solid and dispersed of the kidney, breast, colon, ovarian and
CC   cervical tumours and melanomas) and in particular, tumours of the blood
CC   such as leukaemia. Alternatively, the polypeptides may be used as
CC   antigens in the production of antibodies to IL-12 and to assay for
CC   agonists and antagonists of its activity. The antibodies and antagonists
CC   may be used to inhibit the activity of IL-12. (I) may also be used
CC   diagnostically as a probe which hybridizes to sequences encoding IL-12
CC   and the antibodies may be used to detect the presence of IL-12
CC   polypeptides in samples. They may be used diagnostically to quantitate
CC   the expression of the polypeptide by patients and hence which subjects
CC   may be in need of restorative therapy.
XX
SQ   Sequence 1399 BP; 390 A; 310 C; 333 G; 366 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 1399;
  Best Local Similarity   87.0%;  Pred. No. 29;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db   785 attctcggcaggtggaggtcagc 807
```

```
RESULT  22
AAZ50892/c
ID   AAZ50892 standard; cDNA; 2096 BP.
XX
AC   AAZ50892;
XX
DT   31-MAY-2000  (first entry)
XX
DE   Human receptor-associated protein cDNA from Incyte clone 3407686.
XX
KW   Human receptor-associated protein; HRAP; Incyte clone 3407686;
KW   cytostatic; immunomodulatory; antiinflammatory; cardiant; antianaemic;
KW   antiarteriosclerotic; hepatotropic; antiarthritic antirheumatic;
KW   antiasthmatic; osteopathic; antiallergic; antidiabetic; dermatological;
KW   neuroprotective; diagnosis; treatment; prevention; reproductive disorder;
KW   cardiovascular; cell proliferative; autoimmune; inflammatory; allergy;
KW   gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;
KW   arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;
KW   multiple sclerosis; irritable bowel syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             346..1449
FT                   /*tag=  a
FT                   /product= "HRAP"
XX
PN   WO200008155-A2.
```

```
XX
PD   17-FEB-2000.
XX
PF   06-AUG-1999;   99WO-US17777.
XX
PR   07-AUG-1998;   98US-0160065.
PR   01-SEP-1998;   98US-0098703.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Yue H,  Lal P,  Tang YT,  Gorgone GA,  Guegler KJ;
PI   Corley NC,  Baughn MR;
XX
DR   WPI; 2000-205710/18.
DR   P-PSDB; AAY69990.
XX
PT   New human receptor-associated proteins (HRAP) useful for the diagnosis,
PT   treatment and prevention of cell proliferative, autoimmune,
PT   inflammatory, reproductive, cardiovascular, and gastrointestinal
PT   disorders -
XX
PS   Claim 9; Pages 91-92; 99pp; English.
XX
CC   The present sequence is a cDNA encoding human receptor-associated protein
CC   (HRAP) from Incyte clone 3407686 obtained from PROSTUS08 cDNA library.
CC   This sequence is expressed in reproductive, nervous and
CC   and cardiovascular tissues. HRAP has cytostatic, immunomodulatory,
CC   antiinflammatory, cardiant, antiarteriosclerotic, hepatotropic,
CC   antiarthritic, antirheumatic, osteopathic, antiallergic, antianaemic,
CC   antiasthmatic, antidiabetic, dermatological and neuroprotective
CC   activities. The present sequence is useful in the diagnosis, treatment
CC   and prevention of disorders associated with HRAP expression, especially
CC   cell proliferative, autoimmune/inflammatory, reproductive,
CC   cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
CC   cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
CC   asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
CC   irritable bowel syndrome).
XX
SQ   Sequence 2096 BP; 422 A; 614 C; 572 G; 488 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 2096;
  Best Local Similarity   87.0%;  Pred. No. 30;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      73 GAGCCTCGCCAGGTGCCGGTCAG 51
```

```
RESULT  23
AAA92406
ID   AAA92406 standard; cDNA; 2300 BP.
XX
AC   AAA92406;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Human interleukin 2 encoding p35 and p40 proteins.
XX
KW   Human; interleukin 2; IL-2; p35; p40; immune response; tumour;
KW   infectious disease; bacteria; virus; parasite; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             19..678
FT                   /*tag=  a
FT                   /product= "p35"
FT   CDS             1304..2290
FT                   /*tag=  b
FT                   /product= "p40"
XX
PN   CN1258740-A.
XX
PD   05-JUL-2000.
XX
PF   31-DEC-1998;   98CN-0126748.
XX
PR   31-DEC-1998;   98CN-0126748.
XX
PA   (HUAC-) HUACHEN BIOLOGICAL TECHNOLOGY INST SHANG.
XX
PI   Cao X,  Zhang W,  Ju D;
XX
DR   WPI; 2000-558956/52.
DR   P-PSDB; AAB23795, AAB23796.
XX
PT   Recombinant adenovirus expressing human interleukin-12 and its
PT   preparation and use -
XX
PS   Example 3; Fig 4; 27pp; Chinese.
XX
CC   The present invention describes a recombinant adenovirus and its
CC   preparation and use, as well as medicine composite containing the
CC   recombinant adenovirus. The recombinant adenovirus is used clinically,
CC   has the advantages of low cost, easy use and high efficiency and is
CC   especially suitable for immune treatment of tumour and infectious
CC   diseases caused by bacteria, virus and parasite: The present sequence
CC   represents human interleukin 2 (IL-2) cDNA which encodes the p35 and
CC   p40 proteins, which is used in an example from the present invention.
XX
SQ   Sequence 2300 BP; 598 A; 591 C; 571 G; 540 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 2300;
  Best Local Similarity   87.0%;  Pred. No. 30;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    2046 attctcggcaggtggaggtcagc 2068
```

```
RESULT  24
AAQ23586
ID   AAQ23586 standard; cDNA; 2362 BP.
XX
AC   AAQ23586;
XX
DT   16-SEP-1992  (first entry)
XX
DE   Natural killer cell stimulatory factor 40kD subunit.
XX
KW   NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;
KW   IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag=  a
FT                   /product= pro 40 kD subunit of NKSF
XX
PN   WO9205256-A.
XX
PD   02-APR-1992.
XX
PF   04-SEP-1991;   91WO-US06332.
XX
PR   18-SEP-1990;   90US-0584941.
XX
PA   (GENE-) GENETICS INST INC.
PA   (WIST-) WISTAR INST.
XX
```

PI Clark S C, Hewick R, Kobayashi M, Perussia B, Trinchieri G;
PI Wong G G;
XX
DR WPI; 1992-132124/16.
DR P-PSDB; AAR22769.
XX
PT New natural killer cell stimulatory factor - useful as an
PT immunostimulant for inducing gamma-interferon and GM-CSF prodn.
PT in human peripheral blood lymphocytes
XX
PS Claim 12; Page 23; 79pp; English.
XX
CC NKSF is capable of inducing the production of gamma interferon in
CC human peripheral blood lymphocytes. It has subunits of 40 (AAR22769)
CC and 30-35 kD (AAR23729). It is claimed that NKSF is useful in the
CC treatment of bacterial and viral infections e.g. AIDS.
CC Degenerate probes were designed from tryptic fragments of NKSF
CC and used to screen a cDNA library made from poly A RNA from PdBu
CC induced 8866 cells to identify a cDNA encoding the small subunit
CC of NKSF. cDNA prepared from 8866 poly A RNA was cloned into
CC lambdaZAP and screened with a fragment within pNK-6 (A prior
CC partial clone containing the N-terminal non coding region of the
CC 40kD subunit). The library was then screened with probes designed
CC from tryptic peptides as above, to yield pNK162. This was ligated
CC together with pNK-6 to form pNK40-4 which is believed to contain
CC the full length cDNA for the 40 Kd NKSF subunit.
CC See also AAR22769, AAQ24371, AAR23729.
XX
SQ Sequence 2362 BP; 717 A; 482 C; 543 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 13; Length 2362;
Best Local Similarity 87.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 25
AAT48098
ID AAT48098 standard; cDNA; 2362 BP.
XX
AC AAT48098;
XX
DT 21-MAR-1997 (first entry)
XX
DE Human interleukin-12 40 kD subunit gene.
XX
KW Human; interleukin 12; IL-12; heterodimeric complex; immunogenic; ds;
KW Leishmania; antigen; leishmaniasis; cell-mediated immunity; adjuvant;
KW TH1 helper cell; TH2 helper cell; alum; cytokine; bacterium; cancer;
KW vaccine; cell surface; membrane; glycoprotein; antigen presenting cell.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..1019
FT /*tag= a
FT /product= interleukin-12 40 kD subunit
XX
PN US5571515-A.
XX
PD 05-NOV-1996.
XX
PF 18-APR-1994; 94US-0229282.
XX
PR 17-JUN-1994; 94US-0265087.
PR 18-APR-1994; 94US-0229282.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Scott P, Trinchieri G;
XX
DR WPI; 1996-505347/50.
DR P-PSDB; AAW07398.
XX
PT Immunogenic compsn. to improve cell mediated immune response -
PT contains soluble leishmania antigen and interleukin-12 as adjuvant
XX
PS Disclosure; Column 17-22; 22pp; English.
XX
CC This is the nucleotide sequence encoding the human interleukin (IL)-12
CC 40 kD subunit. IL-12 is a heterodimeric complex composed of the 40 kD
CC subunit and a 30 kD subunit (AAW07399). The complex is used in a novel
CC immunogenic composition comprising a soluble Leishmania antigen with
CC IL-12, for protection against leishmaniasis. The addition of IL-12
CC improves cell-mediated immunity by inducing TH1 helper cells (as opposed
CC to TH2 cells which are induced by alum adjuvant) and does not cause
CC uncontrolled release of other cytokines (in contrast to bacterial
CC adjuvants). IL-12 can also be used as a cancer vaccine by association
CC with the protein B7, a soluble, cell-surface (membrane)-bound
CC glycoprotein which is expressed in antigen presenting cells.
XX
SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 17; Length 2362;
Best Local Similarity 87.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 26
AAV42535
ID AAV42535 standard; cDNA; 2362 BP.
XX
AC AAV42535;
XX
DT 09-NOV-1998 (first entry)
XX
DE Human interleukin-13 40 kDa subunit cDNA.
XX
KW Interleukin-13; IL-13; human; cytokine; adjuvant; vaccine;
KW tetanus toxoid; HIV-1; Mycobacterium tuberculosis; influenza A;
KW hepatitis A; hepatitis B; hepatitis C; leprosy; leishmaniasis;
KW malaria; AIDS; cancer; tumour; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..1019
FT /*tag= a
XX
PN WO9831384-A1.
XX
PD 23-JUL-1998.
XX
PF 22-JAN-1998; 98WO-US01500.
XX
PR 22-JAN-1997; 97US-0036342.
XX
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Montaner LJ;
XX
DR WPI; 1998-413812/35.
DR P-PSDB; AAW62992.
XX
PT Enhancing immune response to antigen using interleukin-13 - useful
PT in vaccine production, especially for antigens not responsive to

```
PT   known adjuvants e.g. Tetanus Toxoid and to treat e.g. HIV or cancer
XX
PS   Disclosure; Page 53-55; 95pp; English.
XX
CC   This cDNA sequence includes an open reading frame coding for the
CC   40 kDa subunit (see AAW62992) of human interleukin-13 (IL-13).  A
CC   claimed method for enhancing the immune response to a selected
CC   antigen in a host comprises administering an effective amount of
CC   IL-13 or a biologically active fragment ('IL-13'), simultaneously
CC   or sequentially with the antigen.  Also claimed are: (1) a method
CC   for enhancing the adjuvant effect of interleukin-12 (IL-12) on an
CC   antigen in a host by similarly administering IL-13 with the antigen
CC   and IL-12; (2) vaccine compositions comprising a selected antigen
CC   and IL-13, optionally with a second adjuvant; and (3) use of IL-13
CC   to prepare medicaments for administration as above.  The method is
CC   used to enhance the immune response of a host to a particular
CC   antigen, especially when the IL-13 is administered in medicaments as
CC   in (2), optionally further comprising IL-12 or functional fragments
CC   (especially in which the antigen is Tetanus Toxoid (TT), influenza
CC   A, HIV-1 p24, HIV-1 nef or Mycobacterium tuberculosis) (claimed).
CC   IL-13 was shown to be an effective adjuvant in both healthy or
CC   pathogen-infected mammals (e.g. HIV-infected humans) (claimed).
CC   The method and vaccines/medicaments of (1) and (2) may be used to
CC   treat viral infections (e.g. HIV, Influenza A, Hepatitis A, B or C
CC   etc.), bacterial infections (e.g. leprosy or tuberculosis),
CC   parasitic infections (e.g. leishmaniasis or malaria) and certain
CC   cancers and solid tumours.  Use of IL-13 may be preferable to use
CC   of other adjuvants since it is relatively stable in vivo (unlike
CC   IL-2), has a direct action on antigen presenting cells (unlike
CC   IL-12), and unlike bacterial adjuvants does not induce
CC   pro-inflammatory cytokines.  IL-13 can also adjuvant antigens not
CC   responsive to known adjuvants (e.g. TT).  IL-13 nucleic acids can
CC   also be used as adjuvants, operably linked to sequences encoding
CC   IL-2, IL-12 and/or the antigen.
XX
SQ   Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 30;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  27
AAV15227
ID   AAV15227 standard; cDNA; 2362 BP.
XX
AC   AAV15227;
XX
DT   27-MAY-1998  (first entry)
XX
DE   Heterodimeric human interleukin 12 subunit 1 encoding cDNA.
XX
KW   Heterodimeric; human; interleukin 12; IL-12; vaccine; antigen;
KW   pathogen; immune response; microorganism; cancer; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag=  a
FT                   /product= "heterodimeric human interleukin 12 subunit"
XX
PN   US5723127-A.
XX
PD   03-MAR-1998.
XX
PF   25-MAR-1996;   96US-0621493.
XX
PR   17-JUN-1994;   94US-0265087.
PR   18-APR-1994;   94US-0229282.
PR   25-MAR-1996;   96US-0621493.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Scott P,  Trinchieri G;
XX
DR   WPI; 1998-205996/18.
DR   P-PSDB; AAW44324.
XX
PT   Enhancing immune response to vaccine - comprises co-administering
PT   interleukin-12 with antigenic composition
XX
PS   Disclosure; Column 17-22; 23pp; English.
XX
CC   The present sequence encodes a heterodimeric human interleukin 12
CC   (IL-12) subunit (putatively the 40 kD subunit) shown in the present
CC   specification. The present invention describes a method for enhancing
CC   the immune response elicited by an antigenic composition. The method
CC   comprises administering IL-12 simultaneously or sequentially with the
CC   antigenic composition. The IL-12 can be used in therapeutic
CC   compositions, e.g. as an adjuvant in vaccines against pathogenic
CC   bacteria or viruses, especially in an amount of 0.1 mu g to 0.5 mg per
CC   25 mu g of antigen, or in cancer vaccines.
XX
SQ   Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 30;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  28
AAZ28862
ID   AAZ28862 standard; cDNA; 2362 BP.
XX
AC   AAZ28862;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human interleukin-12 40 kD subunit gene.
XX
KW   Human; interleukin-12; IL-12; antigen; cancer; immune response; ds.
XX
OS   Homo sapiens.
XX
PN   US5976539-A.
XX
PD   02-NOV-1999.
XX
PF   06-NOV-1997;   97US-0965688.
XX
PR   17-JUN-1994;   94US-0265087.
PR   25-MAR-1996;   96US-0621493.
PR   18-APR-1994;   94US-0229282.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Trinchieri G,  Scott P;
XX
DR   WPI; 2000-021927/02.
DR   P-PSDB; AAY44181.
XX
PT   A new anticancer vaccine  -
XX
```

PS Disclosure; Column 17-22; 23pp; English.
XX
CC This sequence represents the coding region for the human interleukin-12
CC (IL-12) 40 kD subunit (see AAZ28863 for the coding sequence of the 30 kD
CC subunit). IL-12 can be used in a composition with an antigen from a
CC cancer cell or a cancer cell transfected with an antigen, to stimulate
CC an immune response to the anticancer composition, treating cancer.
XX
SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 21; Length 2362;
Best Local Similarity 87.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 29
AAF56280
ID AAF56280 standard; DNA; 2362 BP.
XX
AC AAF56280;
XX
DT 18-APR-2001 (first entry)
XX
DE IL-12 vaccine adjuvant related DNA #1.
XX
KW Il-12; vaccine; adjuvant; immune; virus; ds.
XX
OS Unidentified.
XX
PN US6168923-B1.
XX
PD 02-JAN-2001.
XX
PF 01-MAR-1999; 99US-0260173.
XX
PR 17-JUN-1994; 94US-0265087.
PR 25-MAR-1996; 96US-0621493.
PR 06-NOV-1997; 97US-0965688.
PR 18-APR-1994; 94US-0229282.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Trinchieri G, Scott P;
XX
DR WPI; 2001-122259/13.
XX
PT Use of interleukin 12 or a polynucleotide sequence encoding IL-12 as
PT adjuvant in vaccine compositions comprising an antigen from a pathogen
PT or a cancer cell -
XX
PS Disclosure; Column 17-22; 23pp; English.
XX
CC The present invention relates to a composition of an adjuvanting
CC amount of interleukin (IL)-12 and an antigen from a pathogenic
CC microorganism, designed to elicit a host cell mediated immune
CC response. The invention may be used to generate an immune response
CC against antigens from an intracellular parasite or a virus such as HIV,
CC hepatitis A, hepatitis B, hepatitis C, rabies virus, Herpes viruses,
CC papilloma virus, yellow fever virus, Epstein-Barr virus,
CC respiratory syncytical virus, parvovirus, chikungunya virus,
CC hemorrhagic fever viruses, and Klebsiella.
XX
SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 22; Length 2362;
Best Local Similarity 87.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 30
AAZ40415
ID AAZ40415 standard; DNA; 5686 BP.
XX
AC AAZ40415;
XX
DT 15-FEB-2000 (first entry)
XX
DE Plasmid pIN1143 containing human IL-12 sequence.
XX
KW Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW cancer; intron; ss.
XX
OS Synthetic.
XX
PN WO9947678-A2.
XX
PD 23-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05394.
XX
PR 19-MAR-1998; 98US-0078654.
XX
PA (GENE-) GENEMEDICINE INC.
XX
PI Nordstrom J, Pericle F, Rolland A, Ralston R;
XX
DR WPI; 1999-562116/47.
XX
PT New plasmids containing an interferon-alpha coding sequence, used for
PT the treatment of a mammalian condition or disease, particularly cancer
PT -
XX
PS Disclosure; Fig 2; 137pp; English.
XX
CC The invention relates to a novel plasmid comprising a cytomegalovirus
CC (CMV) promoter transcriptionally linked with an interferon alpha
CC (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC region (UTR). This sequence represents the plasmid pIN1143 which
CC contains the human interleukin 12 (IL-12) gene. The plasmids can be
CC used for treating a mammalian condition or disease, particularly cancer.
XX
SQ Sequence 5686 BP; 1367 A; 1517 C; 1446 G; 1356 T; 0 other;

Query Match 72.8%; Score 18.2; DB 20; Length 5686;
Best Local Similarity 87.0%; Pred. No. 32;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 3961 attctcggcaggtggaggtcagc 3983

RESULT 31
AAZ40023
ID AAZ40023 standard; DNA; 6139 BP.
XX
AC AAZ40023;
XX
DT 14-FEB-2000 (first entry)
XX
DE Interleukin-12 fusion protein, Flexi-12, coding sequence.
XX
KW Interleukin-12; IL-12; fusion protein; IL-12 p35 subunit; B7 protein;

KW IL-12 p40 subunit; gene therapy; tumour; leukaemia; Flexi-12; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US5994104-A.
XX
PD 30-NOV-1999.
XX
PF 08-NOV-1996; 96US-0751767.
XX
PR 08-NOV-1996; 96US-0751767.
XX
PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI Anderson RJ, Prentice HG, MacDonald ID;
XX
DR WPI; 2000-038261/03.
DR P-PSDB; AAY54921.
XX
PT Nucleic acid constructs encoding interleukin-12 fusion proteins useful
PT for treating leukemia and other cancers -
XX
PS Example; Fig 11; 73pp; English.
XX
CC This sequence encodes the human interleukin-12 fusion protein sequence,
CC designated Flexi-12. The invention relates to an isolated nucleic acid
CC construct (I) comprising a region encoding an interleukin-12 (IL-12)
CC fusion protein (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and
CC a linker peptide (joining the subunits)) and a region encoding a B7
CC protein. (I) may be used to produce IL-12 fusion proteins according to
CC standard recombinant DNA methodologies. The fusion proteins may be
CC produced either in vitro in a fermentation culture or in vivo as part of
CC a gene therapy protocol (in this case (I) is used to transform a patients
CC cells, which then secrete the functional polypeptide to supplement the
CC patients own production of IL-12 or to rectify mutations which lead to
CC the expression of inactive polypeptides). The fusion proteins produced in
CC this way may be used to treat any disease which responds to IL-12 such as
CC tumours (both solid and dispersed of the kidney, breast, colon, ovarian
CC and cervical tumours and melanomas) and in particular, tumours of the
CC blood such as leukaemia. Alternatively, the polypeptides may be used as
CC antigens in the production of antibodies to IL-12 and to assay for
CC agonists and antagonists of its activity. The antibodies and antagonists
CC may be used to inhibit the activity of IL-12. (I) may also be used
CC diagnostically as a probe which hybridizes to sequences encoding IL-12
CC and the antibodies may be used to detect the presence of IL-12
CC polypeptides in samples. They may be used diagnostically to quantitate
CC the expression of the polypeptide by patients and hence which subjects
CC may be in need of restorative therapy.
XX
SQ Sequence 6139 BP; 1552 A; 1538 C; 1558 G; 1491 T; 0 other;

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 33;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    3964 attctcggcaggtggaggtcagc 3986
```

RESULT 32
AAC75563/c
ID AAC75563 standard; cDNA; 586 BP.
XX
AC AAC75563;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1118 polynucleotide sequence SEQ ID NO:2235.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41354.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 1649; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;

```
  Query Match             68.8%;  Score 17.2;  DB 21;  Length 586;
  Best Local Similarity   86.4%;  Pred. No. 75;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcag 24
           |||||||||||||||| ||  ||
Db     304 ATCCTCGTCAGGTGCCGGGTAG 283
```

```
RESULT  33
AAC76615/c
ID   AAC76615 standard; cDNA; 5269 BP.
XX
AC   AAC76615;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2170 polynucleotide sequence SEQ ID NO:4339.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42406.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3533-3536; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 5269 BP; 1031 A; 1619 C; 1568 G; 1049 T; 2 other;

  Query Match             67.2%;  Score 16.8;  DB 21;  Length 5269;
  Best Local Similarity   90.0%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcagg 20
           |||||| ||||||| |||||
Db     439 CGATCCACGTCAGGCGCAGG 420


RESULT  34
AAQ75832
ID   AAQ75832 standard; DNA; 34 BP.
XX
AC   AAQ75832;
XX
DT   18-AUG-1995  (first entry)
XX
DE   Sense primer to amplify Non-A Non-B hepatitis virus for analysis.
XX
KW   Non-A Non-B hepatitis virus; structural region; cDNA to genomic RNA;
KW   detection; reagent; anti-Non-A Non-B hepatitis virus antibody;
KW   vaccine; antigen; epitope; diagnosis; ss.
XX
OS   Synthetic.
XX
PN   EP628572-A.
XX
PD   14-DEC-1994.
XX
PF   27-MAY-1994;   94EP-0108256.
XX
PR   28-MAY-1993;   93JP-0126709.
PR   02-MAR-1994;   94JP-0032201.
XX
PA   (ARIM/) ARIMA T.
PA   (EISA ) EISAI CO LTD.
XX
PI   Aoyama M,  Arima T,  Hosoda T,  Iwasaki Y,  Obara T;
PI   Sawada T, Tohmatsu J;
XX
DR   WPI; 1995-015655/03.
XX
PT   New non-A non-B hepatitis virus sub-type - used to develop prods.
PT   for detection, diagnosis, prevention and treatment of non-A non-B
PT   hepatitis.
XX
PS   Example 2; Page 49; 59pp; English.
XX
CC   This primer is based on nucleotides 778-800 of AAQ75818 (a part of
CC   the Non-A Non-B hepatitis virus genome encoding the non-structural
CC   protein). It is used in conjunction with AAQ75833 to amplify nucleotides
CC   800-1326 of AAQ75818.  The nucleotide sequences (see also AAQ75817-19)
CC   were isolated from the plasma of donors in Japan with high s-GTP levels,
CC   and were found to be different from previously reported NANB hepatitis
CC   viruses. The DNA can be used as a reagent for detecting the NANB
CC   hepatitis viral gene. The polypeptides can be used as reagents for
CC   detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
CC   vaccine.
XX
SQ   Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 16;  Length 34;
  Best Local Similarity   82.6%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           |||||||||||||||| | | | |
Db       8 gatcctcgtcaggtgaaaggccg 30
```

```
RESULT  35
AAF28005
ID   AAF28005 standard; cDNA; 220 BP.
XX
AC   AAF28005;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Human TR13 coding sequence fragment SEQ ID NO: 14.
XX
KW   Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW   cancer; autoimmune disease; allergy; inflammatory disease;
KW   graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS   Homo sapiens.
XX
PN   WO200105834-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19343.
XX
PR   16-JUL-1999;   99US-0144087.
PR   18-AUG-1999;   99US-0149450.
PR   20-AUG-1999;   99US-0149712.
PR   10-SEP-1999;   99US-0153089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Young PE;
XX
DR   WPI; 2001-112682/12.
XX
PT   Nucleic acids encoding 2 human tumor necrosis factor receptor
PT   polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT   and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT   hypohidrotic ectodermal dysplasia -
XX
PS   Claim 31; Page 382; 418pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC   useful in the diagnosis and treatment of many diseases, including cancer,
CC   autoimmune diseases, cardiovascular disorders, allergies,
CC   neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC   infections.
XX
SQ   Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 220;
  Best Local Similarity   82.6%;  Pred. No. 1.3e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | ||||  ||||||||||||  ||
Db      59 ccatccgcgtcaggtgcagtcca 81


RESULT  36
AAF28044
ID   AAF28044 standard; cDNA; 220 BP.
XX
AC   AAF28044;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Human TR13 coding sequence fragment SEQ ID NO: 55.
XX
KW   Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW   cancer; autoimmune disease; allergy; inflammatory disease;
KW   graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS   Homo sapiens.
XX
PN   WO200105834-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19343.
XX
PR   16-JUL-1999;   99US-0144087.
PR   18-AUG-1999;   99US-0149450.
PR   20-AUG-1999;   99US-0149712.
PR   10-SEP-1999;   99US-0153089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Young PE;
XX
DR   WPI; 2001-112682/12.
XX
PT   Nucleic acids encoding 2 human tumor necrosis factor receptor
PT   polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT   and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT   hypohidrotic ectodermal dysplasia -
XX
PS   Claim 33; Page 407; 418pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC   useful in the diagnosis and treatment of many diseases, including cancer,
CC   autoimmune diseases, cardiovascular disorders, allergies,
CC   neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC   infections.
XX
SQ   Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 220;
  Best Local Similarity   82.6%;  Pred. No. 1.3e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | ||||  ||||||||||||  ||
Db      59 ccatccgcgtcaggtgcagtcca 81


RESULT  37
AAC98199
ID   AAC98199 standard; cDNA; 625 BP.
XX
AC   AAC98199;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:209.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;.
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
```

PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR P-PSDB; AAB53442.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 1; Page 633; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 625 BP; 139 A; 177 C; 143 G; 156 T; 10 other;

Query Match 66.4%; Score 16.6; DB 21; Length 625;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 cgatcctcgtcaggtgcaggtca 23
          | |||| |||||||||||  ||
Db    464 ccatccgcgtcaggtgcagtcca 486
```

RESULT 38
AAF14300
ID AAF14300 standard; cDNA; 651 BP.
XX
AC AAF14300;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:6823.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2778; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 651 BP; 185 A; 171 C; 162 G; 133 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 651;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 cgatcctcgtcaggtgcaggtca 23
          || |||||||||| | ||||| |
Db    552 cggtcctcgtcagttccaggtga 574
```

RESULT 39
AAF71312
ID AAF71312 standard; DNA; 681 BP.
XX
AC AAF71312;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:349.
XX
KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
KW Brevibacterium; environmental condition; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100842-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00911.
XX

PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032127.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032129.
PR 09-JUL-1999; 99DE-1032226.
PR 14-JUL-1999; 99DE-1032920.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1032930.
PR 14-JUL-1999; 99DE-1032933.
PR 14-JUL-1999; 99DE-1032935.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033002.
PR 14-JUL-1999; 99DE-1033003.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041390.
PR 31-AUG-1999; 99DE-1041391.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061974/07.
DR P-PSDB; AAB79197.
XX
PT New isolated Corynebacterium glutamicum nucleic acid for production or
PT modulation of production of fine chemicals such as amino acids,
PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT or enzymes -
XX
PS Claim 3; Page 597-598; 712pp; English.
XX
CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC C. glutamicum HA genes (I) can be used in vectors for expression in host
CC cells and production of fine chemicals, such as, an organic acid,
CC proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC be modulated. The presence of (I) or HA proteins encoded by then are
CC used for diagnosing the presence or activity of Corynebacterium
CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC used as markers for genetically engineered Corynebacterium or
CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
CC homeostasis in C. glutamicum or help the microorganism to adapt to
CC different environmental conditions.
XX
SQ Sequence 681 BP; 149 A; 164 C; 221 G; 147 T; 0 other;

Query Match 66.4%; Score 16.6; DB 22; Length 681;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcag 24
|||| ||| ||| ||||||| ||
Db 78 gatcgtcgccagctgcaggtgag 100

RESULT 40
AAX18177
ID AAX18177 standard; cDNA; 924 BP.
XX
AC AAX18177;
XX
DT 07-MAY-1999 (first entry)
XX
DE Canine mature CLAF p40 subunit coding sequence.
XX
KW CLAF; canine lymphocyte activation factor; p35 subunit; p40 subunit; dog;
KW viral disease; canine distemper; canine parvovirus; autoimmune disease;
KW canine infectious hepatitis; ds.
XX
OS Canis sp.
XX
PN WO9855511-A1.
XX
PD 10-DEC-1998.
XX
PF 26-MAY-1998; 98WO-JP02295.
XX
PR 03-JUN-1997; 97JP-0161936.
XX
PA (KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
XX
PI Eda Y, Imamura T, Maeda H, Tokiyoshi S;
XX
DR WPI; 1999-070260/06.
DR P-PSDB; AAW74323.
XX
PT Canine cytokine protein which activates canine cytotoxic T
PT lymphocytes - is useful for treatment of viral and autoimmune
PT diseases in dogs
XX
PS Claim 8; Page 32; 51pp; Japanese.
XX
CC This sequence encodes the canine lymphocyte activation factor (CLAF) of
CC the invention. CLAF activates lymphocytes, especially canine cytotoxic
CC T cells. The protein contains a p35 and a p40 subunit. CLAF or antibodies
CC recognising it may be used in the treatment of viral diseases in dogs
CC (such as canine distemper, canine parvovirus and canine infectious
CC hepatitis). The antibodies may also be used in isolation of CLAF from
CC culture of the transformant host cells by affinity chromatography. The
CC p40 homodimer can be used for the treatment of autoimmune diseases in
CC dogs.
XX
SQ Sequence 924 BP; 273 A; 222 C; 234 G; 195 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 924;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| || ||| ||||||||
Db 683 attctcggcacgtggaggtcagc 705

RESULT 41
AAV13816
ID AAV13816 standard; cDNA to mRNA; 990 BP.
XX
AC AAV13816;
XX
DT 14-MAY-1998 (first entry)
XX
DE Canine IL-12 P40 subunit cDNA.
XX
KW Canine; interleukin-12 P40 subunit; IL-12 P40 subunit; antitumour;
KW antiviral; vaccine adjuvant; ds.
XX
OS Canis sp.
XX
FH Key Location/Qualifiers

```
FT   CDS             1..990
FT                   /*tag=  a
FT                   /product= interleukin-12_P40_subunit
XX
PN   JP10036397-A.
XX
PD   10-FEB-1998.
XX
PF   08-NOV-1996;    96JP-0296789.
XX
PR   23-MAY-1996;    96JP-0128104.
PR   08-NOV-1995;    95JP-0289729.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-174914/16.
DR   P-PSDB; AAW41790.
XX
PT   Canine interleukin 12 - comprises P40 and P35 subunits; useful in
PT   veterinary medicine, e.g. antitumour, antivirus and vaccine adjuvant
PT   activities are expected
XX
PS   Claim 8; Page 8; 12pp; Japanese.
XX
CC   The present sequence encodes a canine interleukin-12 (IL-12) P40
CC   subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
CC   of inducing an antiviral activating factor and the expression of
CC   class II MHC molecules in canine tumour cells, stimulating
CC   proliferation of canine blastogenic lymphocytes and activating
CC   canine leukocytes to inhibit canine tumour cells. The canine IL-12
CC   can be used in veterinary medicines, e.g. antitumour, antivirus and
CC   vaccine adjuvant activities are expected.
XX
SQ   Sequence 990 BP; 278 A; 243 C; 251 G; 218 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 19;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 attctcggcacgtggaggtcagc 768


RESULT  42
AAV13817
ID   AAV13817 standard; cDNA to mRNA; 990 BP.
XX
AC   AAV13817;
XX
DT   14-MAY-1998  (first entry)
XX
DE   Canine IL-12 P40 subunit cDNA.
XX
KW   Canine; interleukin-12 P40 subunit; IL-12 P40 subunit; antitumour;
KW   antiviral; vaccine adjuvant; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
FT                   /product= interleukin-12_P40_subunit
XX
PN   JP10036397-A.
XX
PD   10-FEB-1998.
XX
PF   08-NOV-1996;    96JP-0296789.
XX
PR   23-MAY-1996;    96JP-0128104.
PR   08-NOV-1995;    95JP-0289729.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-174914/16.
DR   P-PSDB; AAW41791.
XX
PT   Canine interleukin 12 - comprises P40 and P35 subunits; useful in
PT   veterinary medicine, e.g. antitumour, antivirus and vaccine adjuvant
PT   activities are expected
XX
PS   Claim 10; Pages 10-11; 12pp; Japanese.
XX
CC   The present sequence encodes a canine interleukin-12 (IL-12) P40
CC   subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
CC   of inducing an antiviral activating factor and the expression of
CC   class II MHC molecules in canine tumour cells, stimulating
CC   proliferation of canine blastogenic lymphocytes and activating
CC   canine leukocytes to inhibit canine tumour cells. The canine IL-12
CC   can be used in veterinary medicines, e.g. antitumour, antivirus and
CC   vaccine adjuvant activities are expected.
XX
SQ   Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 19;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 attctcggcacgtggaggtcagc 768


RESULT  43
AAX35615
ID   AAX35615 standard; cDNA to mRNA; 990 BP.
XX
AC   AAX35615;
XX
DT   09-JUL-1999  (first entry)
XX
DE   Nucleic acid encoding canine Interleukin-12 (IL-12).
XX
KW   Interleukin-12; IL-12; dog; cat; immune disease; CaIL12; heterodimer;
KW   tumour; skin disease; infectious disease; allergic disease; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
XX
PN   JP11106350-A.
XX
PD   20-APR-1999.
XX
PF   15-MAY-1998;    98JP-0133345.
XX
PR   07-AUG-1997;    97JP-0213755.
PR   16-MAY-1997;    97JP-0127690.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1999-308068/26.
DR   P-PSDB; AAY02339.
XX
PT   A prevention and treating agent containing interleukin 12 (CaIL12) -
PT   for prevention and treatment of dog and cat immune diseases
XX
PS   Claim 1; Page 12; 16pp; Japanese.
XX
CC   The present sequence encodes canine interleukin-12 (IL-12). The
CC   specification describes a method for the prevention and treatment
```

```
CC   of dog and cat immune diseases. The treatment used an agent comprising
CC   dog IL-12 (CaIL12) proteins to form a heterodimer. The agent is
CC   useful for preventing and treating dog and cat immune diseases,
CC   including tumours, skin diseases, infectious diseases and allergic
CC   diseases.
XX
SQ   Sequence 990 BP; 278 A; 243 C; 251 G; 218 T; 0 other;


  Query Match            66.4%;  Score 16.6;  DB 20;  Length 990;
  Best Local Similarity  82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| || ||| ||||||||
Db    746 attctcggcacgtggaggtcagc 768


RESULT  44
AAX35625
ID   AAX35625 standard; cDNA to mRNA; 990 BP.
XX
AC   AAX35625;
XX
DT   09-JUL-1999  (first entry)
XX
DE   Nucleic acid encoding canine Interleukin-12 (IL-12).
XX
KW   Interleukin-12; IL-12; dog; cat; immune disease; CaIL12; heterodimer;
KW   tumour; skin disease; infectious disease; allergic disease; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
XX
PN   JP11106350-A.
XX
PD   20-APR-1999.
XX
PF   15-MAY-1998;    98JP-0133345.
XX
PR   07-AUG-1997;    97JP-0213755.
PR   16-MAY-1997;    97JP-0127690.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1999-308068/26.
DR   P-PSDB; AAY02342.
XX
PT   A prevention and treating agent containing interleukin 12 (CaIL12) -
PT   for prevention and treatment of dog and cat immune diseases
XX
PS   Claim 1; Page 14-15; 16pp; Japanese.
XX
CC   The present sequence encodes canine interleukin-12 (IL-12). The
CC   specification describes a method for the prevention and treatment
CC   of dog and cat immune diseases. The treatment used an agent comprising
CC   dog IL-12 (CaIL12) proteins to form a heterodimer. The agent is
CC   useful for preventing and treating dog and cat immune diseases,
CC   including tumours, skin diseases, infectious diseases and allergic
CC   diseases.
XX
SQ   Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;


  Query Match            66.4%;  Score 16.6;  DB 20;  Length 990;
  Best Local Similarity  82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| || ||| ||||||||
Db    746 attctcggcacgtggaggtcagc 768


RESULT  45
AAX03162
ID   AAX03162 standard; cDNA to mRNA; 990 BP.
XX
AC   AAX03162;
XX
DT   30-MAR-1999  (first entry)
XX
DE   Nucleic acid encoding canine interleukin 12.
XX
KW   Canine; interleukin 12; IL-12; feline; immunological disease; tumour;
KW   skin disease; viral infection; allergic disease; breast tumour;
KW   oesinophilic granuloma; epidermoid tumour; skin tumour; lipoma;
KW   othematoma; pneumonoedema; skin soft pedicled soft tumour; anal tumour;
KW   otitis externa; dermatitis; eczema; fungal skin disease; pyoderma;
KW   allergic dermatitis; nettle rash; traumatic dermatitis; hair loss;
KW   dog Parvovirus infection; distemper virus; cat plaque virus infection;
KW   feline leukaemia; allergy; pollinosis; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
FT                   /product= interleukin_12
XX
PN   WO9851327-A1.
XX
PD   19-NOV-1998.
XX
PF   07-MAY-1998;    98WO-JP02031.
XX
PR   16-MAY-1997;    97JP-0127690.
XX
PA   (TORA ) TORAY IND INC.
XX
PI   Okano F,  Satoh M,  Yamada K;
XX
DR   WPI; 1999-070100/06.
DR   P-PSDB; AAW84373.
XX
PT   New therapeutic and prophylactic agents  - comprise
PT   genetically-engineered canine interleukin 12, used to treat, e.g.
PT   canine and feline immunological diseases
XX
PS   Claim 1; Page 30-32; 45pp; Japanese.
XX
CC   The present sequence encodes a canine interleukin 12 (IL-12) protein.
CC   The IL-12 protein can be used in therapeutic or prophylactic agents.
CC   The agents can be used to prevent and treat canine and feline
CC   immunological diseases including dog and cat tumours, skin diseases,
CC   viral infections and allergic diseases, especially tumours, breast
CC   tumour, oesinophilic granuloma, epidermoid tumour, skin tumour, lipoma,
CC   othematoma, pneumonoedema, skin soft pedicled soft tumour and anal
CC   tumour; skin diseases, otitis externa, dermatitis, eczema, fungal
CC   diseases of the skin, pyoderma, allergic dermatitis, nettle rash,
CC   traumatic dermatitis and hair loss; infections; dog Parvovirus infection
CC   and distemper virus, cat plaque virus infection and feline leukaemia, and
CC   allergic diseases, e.g. pollinosis.
XX
SQ   Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;


  Query Match            66.4%;  Score 16.6;  DB 20;  Length 990;
  Best Local Similarity  82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| || ||| ||||||||
Db    746 attctcggcacgtggaggtcagc 768
```

Search completed: October 9, 2001, 11:40:03
Job time: 1918 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:59 ; Search time 221.13 Seconds
(without alignments)
21.403 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
	5	18.2	72.8	987	1	US-08-186-529-1	Sequence 1, Appli
	6	18.2	72.8	987	1	US-08-640-386A-1	Sequence 1, Appli
	7	18.2	72.8	987	4	US-08-848-760B-24	Sequence 24, Appl
	8	18.2	72.8	1018	2	US-08-184-009-194	Sequence 194, App
	9	18.2	72.8	1018	2	US-08-458-356-194	Sequence 194, App
	10	18.2	72.8	1399	2	US-08-751-767A-3	Sequence 3, Appli
	11	18.2	72.8	1560	2	US-08-751-767A-11	Sequence 11, Appl
	12	18.2	72.8	1623	2	US-08-751-767A-9	Sequence 9, Appli
	13	18.2	72.8	2362	1	US-08-265-087-1	Sequence 1, Appli
	14	18.2	72.8	2362	1	US-08-621-493-1	Sequence 1, Appli
	15	18.2	72.8	2362	2	US-08-965-688-1	Sequence 1, Appli
	16	18.2	72.8	2362	4	US-09-260-173-1	Sequence 1, Appli
	17	18.2	72.8	6139	2	US-08-751-767A-7	Sequence 7, Appli
	18	16.6	66.4	990	4	US-09-079-984A-1	Sequence 1, Appli
	19	16.6	66.4	990	4	US-09-079-984A-11	Sequence 11, Appl
c	20	16.6	66.4	3449	3	US-09-049-672A-16	Sequence 16, Appl
c	21	16.2	64.8	3509	1	US-08-175-471-6	Sequence 6, Appli
c	22	16.2	64.8	3509	2	US-08-429-054A-12	Sequence 12, Appl
c	23	16.2	64.8	3509	2	US-08-718-777-6	Sequence 6, Appli
c	24	16.2	64.8	3509	3	US-09-078-862-2	Sequence 2, Appli
c	25	16.2	64.8	3509	3	US-09-051-341-6	Sequence 6, Appli
c	26	15.8	63.2	2095	1	US-08-336-408B-3	Sequence 3, Appli
c	27	15.8	63.2	2095	5	PCT-US91-00399-3	Sequence 3, Appli
c	28	15.6	62.4	3336	6	5457037-1	Patent No. 5457037
c	29	15.4	61.6	1722	1	US-08-055-945-1	Sequence 1, Appli
c	30	15.2	60.8	50	2	US-08-850-049-122	Sequence 122, App
c	31	15.2	60.8	50	2	US-08-050-478-122	Sequence 122, App
c	32	15.2	60.8	769	2	US-08-676-279-42	Sequence 42, Appl
c	33	15.2	60.8	1381	4	US-09-426-557-5	Sequence 5, Appli
c	34	15.2	60.8	1463	4	US-09-426-557-1	Sequence 1, Appli
c	35	15.2	60.8	1478	4	US-09-426-557-7	Sequence 7, Appli
c	36	15.2	60.8	1541	4	US-09-426-557-3	Sequence 3, Appli
c	37	15.2	60.8	1704	1	US-08-528-199-2	Sequence 2, Appli
c	38	15.2	60.8	1704	1	US-08-528-199-5	Sequence 5, Appli
c	39	15.2	60.8	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	40	15.2	60.8	53526	3	US-08-658-136-2	Sequence 2, Appli
c	41	15.2	60.8	53577	3	US-08-658-136-1	Sequence 1, Appli
c	42	15	60.0	387	4	US-08-905-223-225	Sequence 225, App
c	43	15	60.0	666	4	US-08-998-416-995	Sequence 995, App
c	44	15	60.0	707	4	US-08-998-416-1093	Sequence 1093, Ap
c	45	15	60.0	1045	1	US-08-317-880-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
;   APPLICANT: Patton, David
;   TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: CIBA-GEIGY Corporation
;     STREET: 7 Skyline Drive
;     CITY: Hawthorne
;     STATE: NY
;     COUNTRY: USA
;     ZIP: 10532
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30B
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/401,068
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/351,970
;     FILING DATE: 08-DEC-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Elmer, James Scott
;     REGISTRATION NUMBER: 36,129
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 919-541-8614
;     TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1041 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1038
;     IDENTIFICATION METHOD: experimental
;     OTHER INFORMATION: /product= "biotin synthase"
;     OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7
```

```
  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0077;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;       LOCATION:  5098..5574
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /codon_start= 5098
;       OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;       OTHER INFORMATION:  /product= "protein"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /gene= "ORFI"
;       OTHER INFORMATION:  /number= 6
;     FEATURE:
;       NAME/KEY:  -10_signal
;       LOCATION:  45..49
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "promoter ptac"
;     FEATURE:
;       NAME/KEY:  -35_signal
;       LOCATION:  23..28
;       OTHER INFORMATION:  /standard_name= "promoter ptac"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  105..119
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  2284..2297
;       OTHER INFORMATION:  /standard_name= "bioC RBS"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  3742..3752
;       OTHER INFORMATION:  /standard_name= "bioA RBS"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  5088..5100
;       OTHER INFORMATION:  /standard_name= "ORFI RBS"
;     FEATURE:
;       NAME/KEY:  terminator
;       LOCATION:  5583..5644
;       OTHER INFORMATION:  /standard_name= "rho-independent
;       OTHER INFORMATION:  transcriptional terminator"
;     FEATURE:
;       NAME/KEY:  stem_loop
;       LOCATION:  5583..5605
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  1..96
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "promoter ptac"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  WO 87/01391 B1
;       FILING DATE:  26-AUG-1986
;       PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0094;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
```

```
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0094;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   5
US-08-186-529-1
; Sequence 1, Application US/08186529
; Patent No. 5573764
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/186,529
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDaniels, Patricia A.
;      REGISTRATION NUMBER:  33,194
;      REFERENCE/DOCKET NUMBER:  GI 5225
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-498-8401
;      TELEFAX:  617-876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  RPMI 8866
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..987
US-08-186-529-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 8.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   6
US-08-640-386A-1
; Sequence 1, Application US/08640386A
; Patent No. 5756085
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/640,386A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI 5225A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-498-8224
;      TELEFAX:  617-876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  RPMI 8866
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..987
US-08-640-386A-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 8.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   7
US-08-848-760B-24
; Sequence 24, Application US/08848760B
; Patent No. 6248721
;   GENERAL INFORMATION:
;        APPLICANT: Chang, Lung-Ji
;        TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines
;        NUMBER OF SEQUENCES: 33
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
;             STREET: 2421 N.W. 41st Street, Suite A-1
;             CITY: Gainesville
;             STATE: Florida
;             COUNTRY: United States of America
```

```
;            ZIP: 32606
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/848,760B
;            FILING DATE: 25-Jan-2001
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/838,702
;            FILING DATE: 09-APR-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: PACE, DORAN R.
;            REGISTRATION NUMBER: 38,261
;            REFERENCE/DOCKET NUMBER: CNG-100C1
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (352) 375-8100
;            TELEFAX: (352) 372-5800
;   INFORMATION FOR SEQ ID NO: 24:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 987 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;            DESCRIPTION: /desc = "DNA"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-08-848-760B-24


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 8.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   8
US-08-184-009-194
; Sequence 194, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-184-009-194


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 8.8;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT   9
US-08-458-356-194
; Sequence 194, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,356
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-458-356-194


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 8.8;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
```

```
              || |||| |||||| ||||||||
Db     774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT  10
US-08-751-767A-3
; Sequence 3, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1399 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  43..1026
US-08-751-767A-3


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1399;
  Best Local Similarity   87.0%;  Pred. No. 9.1;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     785 ATTCTCGGCAGGTGGAGGTCAGC 807


RESULT  11
US-08-751-767A-11
; Sequence 11, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1557
US-08-751-767A-11


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1560;
  Best Local Similarity   87.0%;  Pred. No. 9.2;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     680 ATTCTCGGCAGGTGGAGGTCAGC 702


RESULT  12
US-08-751-767A-9
; Sequence 9, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
```

```
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1623 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1620
US-08-751-767A-9


  Query Match            72.8%;  Score 18.2;  DB 2;  Length 1623;
  Best Local Similarity  87.0%;  Pred. No. 9.3;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  13
US-08-265-087-1
; Sequence 1, Application US/08265087
; Patent No. 5571515
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/265,087
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-265-087-1


  Query Match            72.8%;  Score 18.2;  DB 1;  Length 2362;
  Best Local Similarity  87.0%;  Pred. No. 9.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  14
US-08-621-493-1
; Sequence 1, Application US/08621493
; Patent No. 5723127
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/621,493
;      FILING DATE:  25-MAR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-621-493-1


  Query Match            72.8%;  Score 18.2;  DB 1;  Length 2362;
  Best Local Similarity  87.0%;  Pred. No. 9.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  15
US-08-965-688-1
; Sequence 1, Application US/08965688
; Patent No. 5976539
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/965,688
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/621,493
;      FILING DATE:  25-MAR-1996
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-965-688-1


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 9.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  16
US-09-260-173-1
; Sequence 1, Application US/09260173
; Patent No. 6168923
;  GENERAL INFORMATION:
;       APPLICANT: Scott, Phillip
;                  Trinchieri, Giorgio
;       TITLE OF INVENTION: Compositions and Methods for Use of
;                           IL-12 as an Adjuvant
;       NUMBER OF SEQUENCES: 4
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Center, PO Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/260,173
;            FILING DATE: 01-Mar-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/965,688
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: 08/265,087
;            FILING DATE: 17-JUN-1994
;            APPLICATION NUMBER: US 08/229,282
;            FILING DATE: 18-APR-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: WST51AUSA
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9206
;            TELEFAX: 215-540-5818
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2362 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  33..1016
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-260-173-1


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 9.7;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  17
US-08-751-767A-7
; Sequence 7, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
```

```
;       FILING DATE:  08-NOV-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SADOFF, B.J.
;       REGISTRATION NUMBER:  36,663
;       REFERENCE/DOCKET NUMBER:  117-221
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6139 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3222..4841
US-08-751-767A-7


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    3964 ATTCTCGGCAGGTGGAGGTCAGC 3986


RESULT  18
US-09-079-984A-1
; Sequence 1, Application US/09079984A
; Patent No. 6231850
;  GENERAL INFORMATION:
;    APPLICANT:  Okano, Fumiyoshi, Satoh, Massahiro,
;    APPLICANT:  Yamada, Katsushige
;    TITLE OF INVENTION:  Canine interleukin 12, a production method
;    TITLE OF INVENTION:  thereof, an immune disease treatment method and preventive
;    TITLE OF INVENTION:  method using it
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Miller & Christenbury Intellectual Property
;      ADDRESSEE:  Department of Schnader, Harrison, Segal and Lewis, LLP
;      STREET:  1600 Market Street, 39th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,984A
;      FILING DATE:  15-MAY-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Austin R. Miller
;      REGISTRATION NUMBER:  16,602
;      REFERENCE/DOCKET NUMBER:  1051-98
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-1810
;      TELEFAX:  (215) 568-6946
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;    FEATURE:
;      NAME/KEY:  Canine IL12
;      LOCATION:  1 to 987
;      IDENTIFICATION METHOD:  Similarity
US-09-079-984A-1


  Query Match             66.4%;  Score 16.6;  DB 4;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 ATTCTCGGCACGTGGAGGTCAGC 768


RESULT  19
US-09-079-984A-11
; Sequence 11, Application US/09079984A
; Patent No. 6231850
;  GENERAL INFORMATION:
;    APPLICANT:  Okano, Fumiyoshi, Satoh, Massahiro,
;    APPLICANT:  Yamada, Katsushige
;    TITLE OF INVENTION:  Canine interleukin 12, a production method
;    TITLE OF INVENTION:  thereof, an immune disease treatment method and preventive
;    TITLE OF INVENTION:  method using it
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Miller & Christenbury Intellectual Property
;      ADDRESSEE:  Department of Schnader, Harrison, Segal and Lewis, LLP
;      STREET:  1600 Market Street, 39th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,984A
;      FILING DATE:  15-MAY-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Austin R. Miller
;      REGISTRATION NUMBER:  16,602
;      REFERENCE/DOCKET NUMBER:  1051-98
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-1810
;      TELEFAX:  (215) 568-6946
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;    FEATURE:
;      NAME/KEY:  Canine IL12
;      LOCATION:  1 to 987
;      IDENTIFICATION METHOD:  Similarity
US-09-079-984A-11


  Query Match             66.4%;  Score 16.6;  DB 4;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
```

```
Db    746 ATTCTCGGCACGTGGAGGTCAGC 768


RESULT  20
US-09-049-672A-16/c
; Sequence 16, Application US/09049672A
; Patent No. 6135941
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Tang, Y. Tom
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/049,672A
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0497 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3449 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BLADNOT04
;      CLONE:  1320068
US-09-049-672A-16


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 3449;
  Best Local Similarity   82.6%;  Pred. No. 53;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          | ||||||||||   ||||||||
Db    947 GGTCCTCGTCAGAGCCAGGTCAG 925


RESULT  21
US-08-175-471-6/c
; Sequence 6, Application US/08175471
; Patent No. 5665892
;  GENERAL INFORMATION:
;    APPLICANT:  Van Assche, C.
;    APPLICANT:  Lando, D.
;    APPLICANT:  Bruneau, J.M.
;    APPLICANT:  Voelker, T.
;    TITLE OF INVENTION:  Sucrase Phosphate Synthase (SPS), Its
;    TITLE OF INVENTION:  Process For Preparation, Its cDNA, and Utilization of cDNA
;    TITLE OF INVENTION:  Modify The Expression Of SPS In The Plant Cells
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.0 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.7
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/175,471
;      FILING DATE:  27-DEC-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/672,646
;      FILING DATE:  18 MAR 1991
;      APPLICATION NUMBER:  EP 90402084.9
;      FILING DATE:  20 JUL 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elizabeth Lassen
;      REGISTRATION NUMBER:  31,845
;      NAME:  Donna E. Scherer
;      REGISTRATION NUMBER:  34,719
;      REFERENCE/DOCKET NUMBER:  CGNE 72-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (916) 753-6313
;      TELEFAX:  (916) 753-1510
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-175-471-6


  Query Match             64.8%;  Score 16.2;  DB 1;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 81;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  22
US-08-429-054A-12/c
; Sequence 12, Application US/08429054A
; Patent No. 5917126
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
;    APPLICANT:  JEAN; VOELKER, TONI; GERVAIS, MONICA
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHETASE (SPS),
;    TITLE OF INVENTION:  A PREPARATION METHOD AND cDNA THEREFOR, AND USE OF THE
;    TITLE OF INVENTION:  cDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN AND MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/429,054A
;      FILING DATE:  26-APR-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US  842,337
;      FILING DATE:  20-March-1992
;      APPLICATION NUMBER:  PCT/FR 91/00593
;      FILING DATE:  18-July-1991
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  French 90402094.9
;      FILING DATE:  20-July-1990
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles A. Muserlian
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 661-8000
;      TELEFAX:  (212) 661-8002
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Unknown
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  cDNA
US-08-429-054A-12


  Query Match             64.8%;  Score 16.2;  DB 2;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 81;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  23
US-08-718-777-6/c
; Sequence 6, Application US/08718777
; Patent No. 5981852
;  GENERAL INFORMATION:
;    APPLICANT:  Van Assche, C.
;    APPLICANT:  Lando, D.
;    APPLICANT:  Bruneau, J. M.
;    APPLICANT:  Voelker, T.
;    APPLICANT:  Gervais, M.
;    TITLE OF INVENTION:  MODIFICATION OF SUCROSE
;    TITLE OF INVENTION:  PHOSPHATE
;    TITLE OF INVENTION:  SYNTHASE IN PLANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Barbara Rae-Venter
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,777
;      FILING DATE:  NOT YET ASSIGNED
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/175,471
;      FILING DATE:  27-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.072.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..3315
US-08-718-777-6


  Query Match             64.8%;  Score 16.2;  DB 2;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 81;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  24
US-09-078-862-2/c
; Sequence 2, Application US/09078862
; Patent No. 6091003
;  GENERAL INFORMATION:
;    APPLICANT:  Nan, Guo-Ling
;    APPLICANT:  Nagai, Chifumi
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR GENETIC
;    TITLE OF INVENTION:  TRANSFORMATION OF PINEAPPLE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,862
;      FILING DATE:  14-MAY-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  UH-03321
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
US-09-078-862-2


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 81;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  25
US-09-051-341-6/c
; Sequence 6, Application US/09051341
; Patent No. 6124528
;  GENERAL INFORMATION:
;    APPLICANT:  Shewmaker, C. K.
;    TITLE OF INVENTION:  MODIFICATION OF SOLUBLE SOLIDS USING
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rae-Venter Law Group, P.C.
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/051,341
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17351
;      FILING DATE:  25-OCT-1996
;      APPLICATION NUMBER:  US 08/549,016
;      FILING DATE:  27-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/372,200
;      FILING DATE:  12-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter, Ph.D.,
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.110.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..3315
US-09-051-341-6


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 81;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  26
US-08-336-408B-3/c
; Sequence 3, Application US/08336408B
; Patent No. 5723329
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MANGELSDORF, DAVID J.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 SOUTH FLOWER STREET, SUITE 2000
;      CITY:  LOS ANGELES
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,408B
;      FILING DATE:  08-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/933,453
;      FILING DATE:  21-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US91/00399
;      FILING DATE:  22-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REITER, STEPHEN E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9851
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  178..1581
US-08-336-408B-3


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2095;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 ctcgtcaggtgcaggtcag 24
          |||| |||||| |||||||
Db    709 CTCGGCAGGTGTAGGTCAG 691


RESULT  27
PCT-US91-00399-3/c
; Sequence 3, Application PC/TUS9100399
;  GENERAL INFORMATION:
;    APPLICANT:  Mangelsdorf, Dr., David J.
;    APPLICANT:  Evans Dr., Ronald M.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
```

```
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITCH, EVEN, TABIN & FLANNERY
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00399
;      FILING DATE:  19910122
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Mr., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  50852
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 552-1311
;      TELEFAX:  (619) 552-0095
;      TELEX:  20 6566 PATLAW CGO
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  RXR MOUSE ALPHA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(178..1581)
PCT-US91-00399-3


  Query Match             63.2%;  Score 15.8;  DB 5;  Length 2095;
  Best Local Similarity   89.5%;  Pred. No. 1.1e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 ctcgtcaggtgcaggtcag 24
           |||| |||||| |||||||
Db     709 CTCGGCAGGTGTAGGTCAG 691


RESULT  28
5457037-1/c
;Patent No. 5457037
;    APPLICANT: TOGNONI, ANGELO;CARRERA, PAOLO;CAMERINI, BARBARA;
;GALLI, GIULIANO;LUCCHESE, GIUSEPPE;GRANDI, GUIDO;DI GENNARO, CARLO
;    TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
;    NUMBER OF SEQUENCES: 7
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/1,797
;      FILING DATE: 08-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 749,621
;      FILING DATE: 19-AUG-1991
;      APPLICATION NUMBER: 224,114
;      FILING DATE: 25-JUL-1988
;SEQ ID NO:1:
;      LENGTH: 3336
5457037-1


  Query Match             62.4%;  Score 15.6;  DB 6;  Length 3336;
  Best Local Similarity   81.8%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtc 22
           ||||| ||  |||||||| |||
Db      81 CGATCATCCACAGGTGCACGTC 60


RESULT  29
US-08-055-945-1/c
; Sequence 1, Application US/08055945
; Patent No. 5462855
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Rast, Hans George;
;    APPLICANT:  L bberding, Antonius; and Kanne, Reinhard
;    TITLE OF INVENTION:  A METHOD FOR THE DETECTION AND
;    TITLE OF INVENTION:  QUANTITATIVE DETERMINATION OF
;    TITLE OF INVENTION:  NITROSOMONAS STRAINS IN
;    TITLE OF INVENTION:  WASTEWATERS OR SOILS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  1140 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 1.2 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/055,945
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/692,745
;      FILING DATE:  No. 5462855e
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8194-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 391-0520
;      TELEFAX:  (212) 382-0949
;      TELEX:  423092 NYP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 bp
;      TYPE:  Nucleotide
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Genomic DNA
US-08-055-945-1


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 1722;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||   |||||| | |||| |
Db     970 CGATCCTGAACAGGTGGATGTCAAC 946


RESULT  30
US-08-850-049-122/c
; Sequence 122, Application US/08850049
; Patent No. 5965726
```

```
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  METHOD OF ELIMINATING
;    TITLE OF INVENTION:  INHIBITORY/INSTABILITY REGIONS OF mRNA
;    NUMBER OF SEQUENCES:  130
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,049
;      FILING DATE:  02-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,478
;      FILING DATE:  26-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02908
;      FILING DATE:  29-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/858,747
;      FILING DATE:  27-MAR-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MORRY, MARY J.
;      REGISTRATION NUMBER:  34,398
;      REFERENCE/DOCKET NUMBER:  2026-4006US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  122:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
US-08-850-049-122


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 50;
  Best Local Similarity   85.0%;  Pred. No. 1.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcagg 20
           ||||||||||||  |  ||||
Db      43 CGATCCTCGTCACGATCAGG 24


RESULT  31
US-08-050-478-122/c
; Sequence 122, Application US/08050478
; Patent No. 5972596
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  METHOD OF ELIMINATING
;    TITLE OF INVENTION:  INHIBITORY/INSTABILITY REGIONS OF mRNA
;    NUMBER OF SEQUENCES:  130
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,478
;      FILING DATE:  26-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02908
;      FILING DATE:  29-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/858,747
;      FILING DATE:  27-MAR-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MORRY, MARY J.
;      REGISTRATION NUMBER:  34,398
;      REFERENCE/DOCKET NUMBER:  2026-4006US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  122:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
US-08-050-478-122


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 50;
  Best Local Similarity   85.0%;  Pred. No. 1.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcagg 20
           ||||||||||||  |  ||||
Db      43 CGATCCTCGTCACGATCAGG 24


RESULT  32
US-08-676-279-42/c
; Sequence 42, Application US/08676279
; Patent No. 5869247
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MACROPHAGE NUCLEOTIDE SEQUENCE
;    NUMBER OF SEQUENCES:  63
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,279
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB95/00095
;      APPLICATION NUMBER:  GB 9400929.7
;      FILING DATE:  19-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9422021.7
;      FILING DATE:  31-OCT-1994
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  769 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-676-279-42


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 769;
  Best Local Similarity   85.0%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      6 ctcgtcaggtgcaggtcagc 25
          || | ||||| |||||||||
Db    253 CTGGGCAGGTTCAGGTCAGC 234


RESULT  33
US-09-426-557-5/c
; Sequence 5, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 1381
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (37)...(1173)
US-09-426-557-5


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1381;
  Best Local Similarity   85.0%;  Pred. No. 2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtca 23
          ||||| |||||||||||  ||
Db    947 TCCTCATCAGGTGCAGTCCA 928


RESULT  34
US-09-426-557-1/c
; Sequence 1, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1463
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (85)...(1221)
US-09-426-557-1


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1463;
  Best Local Similarity   85.0%;  Pred. No. 2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtca 23
          ||||| |||||||||||  ||
Db    995 TCCTCATCAGGTGCAGTCCA 976


RESULT  35
US-09-426-557-7/c
; Sequence 7, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1478
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (97)...(1233)
US-09-426-557-7


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1478;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtca 23
          ||||| |||||||||||  ||
Db   1007 TCCTCATCAGGTGCAGTCCA 988


RESULT  36
US-09-426-557-3/c
; Sequence 3, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1541
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (79)...(1215)
US-09-426-557-3


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1541;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      4 tcctcgtcaggtgcaggtca 23
          ||||| |||||||||||  ||
Db    989 TCCTCATCAGGTGCAGTCCA 970


RESULT  37
US-08-528-199-2/c
; Sequence 2, Application US/08528199
; Patent No. 5763228
;  GENERAL INFORMATION:
;    APPLICANT:  KUBOTA, Michio
;    APPLICANT:  TSUSAKI, Keiji
;    APPLICANT:  SUGIMOTO, Toshiyuki
;    TITLE OF INVENTION:  RECOMBINANT ENZYME FOR CONVERTING
;    TITLE OF INVENTION:  MALTOSE INTO TREHALOSE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,199
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 156399/1994
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  KUBOTA=6A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-528-199-2


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 1704;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggt 21
          |||| ||| |||||||||  |
Db   1293 GATCGTCGGCAGGTGCAGCT 1274


RESULT  38
US-08-528-199-5/c
; Sequence 5, Application US/08528199
; Patent No. 5763228
;  GENERAL INFORMATION:
;    APPLICANT:  KUBOTA, Michio
;    APPLICANT:  TSUSAKI, Keiji
;    APPLICANT:  SUGIMOTO, Toshiyuki
;    TITLE OF INVENTION:  RECOMBINANT ENZYME FOR CONVERTING
;    TITLE OF INVENTION:  MALTOSE INTO TREHALOSE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,199
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 156399/1994
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  KUBOTA=6A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1704
US-08-528-199-5


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 1704;
  Best Local Similarity   85.0%;  Pred. No. 2.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggt 21
          |||| ||| |||||||||  |
Db   1293 GATCGTCGGCAGGTGCAGCT 1274


RESULT  39
US-08-323-443B-1/c
; Sequence 1, Application US/08323443B
; Patent No. 5654170
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W.
;    APPLICANT:  LANDES, GREGORY M.
;    APPLICANT:  BURN, TIMOTHY C.
;    APPLICANT:  CONNORS, TIMOTHY D.
;    APPLICANT:  DACKOWSKI, WILLIAM R.
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  8
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,443B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  0372/0A462
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  PKD1 GENOMIC
US-08-323-443B-1


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 31571;
  Best Local Similarity   85.0%;  Pred. No. 2.9e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 cctcgtcaggtgcaggtcag 24
           ||||| |||| ||||| |||
Db   13451 CCTCGGCAGGGGCAGGACAG 13432


RESULT  40
US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-2


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 53526;
  Best Local Similarity   85.0%;  Pred. No. 3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 cctcgtcaggtgcaggtcag 24
           ||||| |||| ||||| |||
Db   14436 CCTCGGCAGGGGCAGGACAG 14417


RESULT  41
US-08-658-136-1/c
; Sequence 1, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53577 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
US-08-658-136-1


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 53577;
  Best Local Similarity   85.0%;  Pred. No. 3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 cctcgtcaggtgcaggtcag 24
           ||||| |||| ||||| |||
Db   14435 CCTCGGCAGGGGCAGGACAG 14416


RESULT  42
US-08-905-223-225/c
; Sequence 225, Application US/08905223
; Patent No. 6222029
;  GENERAL INFORMATION:
;    APPLICANT:  Edwards, Jean-Baptiste D.
;    APPLICANT:  Duelert, Aymeric
;    APPLICANT:  Lacroix, Bruno
;    TITLE OF INVENTION:  5' ESTs FOR SECRETED PROTEINS
;    NUMBER OF SEQUENCES:  503
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  501 West Broadway
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92101-3505
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Win95
;      SOFTWARE:  Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/905,223
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  225:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  CDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo Sapiens
;      TISSUE TYPE:  Brain
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(75..325)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 82..332
;      OTHER INFORMATION:  id AA004751
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(88..255)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 97
;      OTHER INFORMATION:  region 153..320
;      OTHER INFORMATION:  id N27443
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(18..105)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 94
;      OTHER INFORMATION:  region 304..391
;      OTHER INFORMATION:  id N27443
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(258..325)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 81..148
;      OTHER INFORMATION:  id N27443
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(22..325)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 80..383
;      OTHER INFORMATION:  id AA015608
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(78..253)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 99
;      OTHER INFORMATION:  region 165..340
;      OTHER INFORMATION:  id H09727
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(253..285)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 96
;      OTHER INFORMATION:  region 132..164
;      OTHER INFORMATION:  id H09727
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(49..276)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 97
;      OTHER INFORMATION:  region 133..360
;      OTHER INFORMATION:  id AA027099
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  other
;      LOCATION:  complement(269..325)
;      IDENTIFICATION METHOD:  blastn
;      OTHER INFORMATION:  identity 100
;      OTHER INFORMATION:  region 83..139
;      OTHER INFORMATION:  id AA027099
;      OTHER INFORMATION:  est
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  139..369
;      IDENTIFICATION METHOD:  Von Heijne matrix
;      OTHER INFORMATION:  score 5.3
;      OTHER INFORMATION:  seq TWVFTCLVFFCFG/LS
US-08-905-223-225


  Query Match             60.0%;  Score 15;  DB 4;  Length 387;
  Best Local Similarity   78.3%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           | ||| || ||||||| || |||
Db     274 ACCCTAGTGAGGTGCATGTGAGC 252
```

```
RESULT  43
US-08-998-416-995/c
; Sequence 995, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  995:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1611RP
US-08-998-416-995


  Query Match             60.0%;  Score 15;  DB 4;  Length 666;
  Best Local Similarity   78.3%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||  ||||||||||   |||
Db     161 GATCGCCGTCAGGTGCAATCCAG 139


RESULT  44
US-08-998-416-1093/c
; Sequence 1093, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1093:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1667RP
US-08-998-416-1093


  Query Match             60.0%;  Score 15;  DB 4;  Length 707;
  Best Local Similarity   78.3%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||  ||||||||||   |||
Db     161 GATCGCCGTCAGGTGCAATCCAG 139


RESULT  45
US-08-317-880-3/c
; Sequence 3, Application US/08317880
; Patent No. 5593862
;  GENERAL INFORMATION:
;    APPLICANT:  Hall, Linda M.
;    TITLE OF INVENTION:  Proteins for Cation Channel Expression
;    TITLE OF INVENTION:  and Function
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/317,880
;       FILING DATE:  04-OCT-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Rogalskyj, Peter
;       REGISTRATION NUMBER:  38,601
;       REFERENCE/DOCKET NUMBER:  19226/610 (R-5261)
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (716) 263-1634
;       TELEFAX:  (716) 263-1600
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1045 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-317-880-3


  Query Match             60.0%;  Score 15;  DB 1;  Length 1045;
  Best Local Similarity   78.3%;  Pred. No. 2.4e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | ||||| | || |||||| |||
Db     158 CTATCCTTGCCACGTGCAGTTCA 136


Search completed: October  9, 2001, 11:40:04
Job time: 1919 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:49 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	18.8	75.2	543	239	AZ168021	AZ168021 SP_0103_B
c 2	18.2	72.8	300	107	AU100606	AU100606 AU100606
c 3	18.2	72.8	354	121	AW826993	AW826993 fk55g12.y
c 4	18.2	72.8	371	112	AW143033	AW143033 EST293328
c 5	18.2	72.8	461	234	AQ846280	AQ846280 LMAJFV1_1
6	18.2	72.8	872	220	CNS026TZ	AL183824 Tetraodon
c 7	18.2	72.8	883	106	AL556337	AL556337 AL556337
c 8	18.2	72.8	904	172	BG023733	BG023733 602303347
c 9	18.2	72.8	917	106	AL580709	AL580709 AL580709
c 10	18.2	72.8	1095	172	BF979422	BF979422 602288251
c 11	17.8	71.2	494	236	AQ982558	AQ982558 RPCI-23-2
12	17.8	71.2	536	165	BE234470	BE234470 141586 MA
13	17.8	71.2	777	106	AL538687	AL538687 AL538687
14	17.8	71.2	1032	221	CNS044UQ	AL274571 Tetraodon
15	17.6	70.4	339	9	AA594461	AA594461 nl94c07.s
c 16	17.6	70.4	360	147	BF318689	BF318689 uy05a01.y
17	17.6	70.4	369	151	BF659032	BF659032 maa89h11.
18	17.6	70.4	398	141	BE860402	BE860402 UI-M-AN1-
c 19	17.6	70.4	413	10	AA644735	AA644735 M-EST008
c 20	17.6	70.4	413	141	BE862745	BE862745 UI-M-BH0-
c 21	17.6	70.4	424	225	AQ172663	AQ172663 HS_3194_A
c 22	17.6	70.4	429	227	AQ371411	AQ371411 HS_5043_A
23	17.6	70.4	463	141	BE850719	BE850719 ux01a08.y
c 24	17.6	70.4	495	24	AI789420	AI789420 uk99a06.y
25	17.6	70.4	508	17	AI180705	AI180705 ub91b01.r
26	17.6	70.4	618	8	AA536738	AA536738 vj88c01.r
27	17.6	70.4	628	138	BE653574	BE653574 UI-M-AM1-
c 28	17.6	70.4	641	243	AZ426129	AZ426129 1M0206H03
c 29	17.6	70.4	721	165	BE262905	BE262905 601153594
c 30	17.6	70.4	833	154	BG478947	BG478947 602525709
c 31	17.6	70.4	893	169	BF794272	BF794272 602255473
32	17.6	70.4	897	220	CNS02FRJ	AL195400 Tetraodon
c 33	17.6	70.4	929	147	BF313136	BF313136 601897654
34	17.6	70.4	942	219	CNS00L1P	AL067857 Drosophil
c 35	17.6	70.4	982	219	CNS006PE	AL065656 Drosophil
c 36	17.6	70.4	985	219	CNS00HWP	AL073836 Drosophil
c 37	17.6	70.4	1002	171	BF965985	BF965985 602277168
c 38	17.6	70.4	1016	219	CNS00HQI	AL074118 Drosophil
c 39	17.4	69.6	1126	140	BE795069	BE795069 601592885
c 40	17.4	69.6	1298	152	BG328084	BG328084 602427134
41	17.2	68.8	108	115	AW371554	AW371554 QV3-BT029
c 42	17.2	68.8	112	10	AA651689	AA651689 nn47b06.r
c 43	17.2	68.8	235	3	AA197041	AA197041 zq10c11.r
44	17.2	68.8	236	147	BF349456	BF349456 RC2-HT014
c 45	17.2	68.8	255	159	N83353	N83353 EC0312F Hum

ALIGNMENTS

```
RESULT   1
AZ168021/c
LOCUS       AZ168021      543 bp    DNA             GSS       29-AUG-2000
DEFINITION  SP_0103_B1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=103 Col=9 Row=L, DNA sequence.
ACCESSION   AZ168021
VERSION     AZ168021.1  GI:8338389
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 103  row: L  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=103 Col=9 Row=L"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      149 a    116 c    116 g    157 t      5 others
ORIGIN


  Query Match             75.2%;  Score 18.8;  DB 239;  Length 543;
  Best Local Similarity   90.9%;  Pred. No. 2.6e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcag 24
           |||||||||||||||| ||| ||
Db     115 ATCCTCGTCAGGTGCTGGTAAG 94


RESULT   2
AU100606/c
LOCUS       AU100606      300 bp    mRNA            EST       05-APR-2001
DEFINITION  AU100606 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            y79a3N51 similar to Homo sapiens mRNA for cystinosin, mRNA
            sequence.
ACCESSION   AU100606
VERSION     AU100606.1  GI:13551735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
  TITLE     In silico mapping of the 5'-ends of human mRNAs using full-length
            enriched and 5'-end enriched cDNA libraries constructed by
            Oligo-capping method
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="y79a3N51"
```

```
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       51 a     93 c     97 g     51 t      8 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 107;  Length 300;
  Best Local Similarity   87.0%;  Pred. No. 4.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     62 GAGCCTCGCCAGGTGCCGGTCAG 40


RESULT   3
AW826993/c
LOCUS       AW826993      354 bp    mRNA            EST       17-MAY-2000
DEFINITION  fk55g12.y1 Zebrafish adult retina cDNA Danio rerio cDNA 5' similar
            to TR:Q15846 Q15846 HYPOTHETICAL 54.2 KD PROTEIN PRECURSOR. ;, mRNA
            sequence.
ACCESSION   AW826993
VERSION     AW826993.1  GI:7920070
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT       83 a    102 c     81 g     87 t      1 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 121;  Length 354;
  Best Local Similarity   83.3%;  Pred. No. 4.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          |||||||| ||  ||||||||||
Db    235 CGATCCTCACCACNTGCAGGTCAG 212


RESULT   4
AW143033/c
LOCUS       AW143033      371 bp    mRNA            EST       30-OCT-1999
DEFINITION  EST293328 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RGIBC08 5' end, mRNA sequence.
ACCESSION   AW143033
VERSION     AW143033.1  GI:6162932
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST please check the TIGR Rat Gene
            Index (http://www.tigr.org/tdb/rgi/rgi.html). To order a clone
            contact the ATCC (http://www.atcc.org/atcc.html).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RGIBC08"
                     /clone_lib="Normalized rat kidney, Bento Soares"
                     /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT       82 a    117 c    103 g     69 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 112;  Length 371;
  Best Local Similarity   87.0%;  Pred. No. 4.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          | ||||| |||||||||||| |||
Db    212 GCTCCTCCTCAGGTGCAGGGCAG 190


RESULT   5
AQ846280/c
LOCUS       AQ846280      461 bp    DNA             GSS       13-APR-2001
DEFINITION  LMAJFV1_lm17b03.y1 Leishmania major FV1 random genomic library
            Leishmania major genomic clone LMAJFV1_lm17b03 5', DNA sequence.
ACCESSION   AQ846280
VERSION     AQ846280.1  GI:6050928
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
            Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
            Blistain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko
            ,I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
            Harvey,N., McCann,R., Tsagareishvili,R., Williams,T., Jackson,Y.,
            Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
  TITLE     A survey of the Leishmania major Friedlin strain V1 genome by
```

```
            shotgun sequencing: a resource for DNA
  JOURNAL   Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
COMMENT     Other_GSSs: lm17b03.x1
            Contact: Akopyants, NS / Beverley, SM
            WashU Leishmania Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library construction: Natalia S. Akopyants, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             If using this information please cite:
             N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
            Friedlin strain V1 genome by shotgun sequencing' and the Washington
            University Genome Sequencing Center For information on obtaining
            clone material please contact: Natalia S. Akopyants Ph.D.
            (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
            (beverley@borcim.wustl.edu)
            Seq primer: -40RP from Gibco
            Class: shotgun
            High quality sequence stop: 373.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Leishmania major"
                     /strain="Friedlin strain V1"
                     /db_xref="taxon:5664"
                     /clone="LMAJFV1_lm17b03"
                     /clone_lib="Leishmania major FV1 random genomic library"
                     /lab_host="TOP10 (Invitrogen)"
                     /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
                     Genomic DNA was isolated from stationary phase cells. For
                     this library, DNA was sheared to give a tight size
                     distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
                     polymerase, dephosphorylated with Shrimp Alkaline
                     Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT      103 a    131 c    139 g     88 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 234;  Length 461;
  Best Local Similarity   87.0%;  Pred. No. 4.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          |||| | ||||| ||||||||||
Db     83 GATCTTGGTCAGCTGCAGGTCAG 61


RESULT   6
CNS026TZ
LOCUS       CNS026TZ      872 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            241J15 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL183824
VERSION     AL183824.1  GI:7821928
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 872)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 872)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="241J15"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG241CE08SP1-end :
                     PUC-Ori"
BASE COUNT      139 a    301 c    254 g    166 t     12 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 220;  Length 872;
  Best Local Similarity   87.0%;  Pred. No. 5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtca 23
          || |||||  |||||||||||||
Db    715 CGCTCCTCTGCAGGTGCAGGTCA 737


RESULT   7
AL556337/c
LOCUS       AL556337      883 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL556337 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK001YI14 5
            prime, mRNA sequence.
ACCESSION   AL556337
VERSION     AL556337.1  GI:12898922
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK001YI14"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      193 a    237 c    228 g    224 t      1 others
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 106;  Length 883;
  Best Local Similarity   87.0%;  Pred. No. 5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     39 GAGCCTCGCCAGGTGCCGGTCAG 17


RESULT   8
BG023733/c
LOCUS       BG023733      904 bp    mRNA            EST       24-JAN-2001
DEFINITION  602303347F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4395058 5',
            mRNA sequence.
ACCESSION   BG023733
VERSION     BG023733.1  GI:12408790
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 904)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10091  row: m  column: 11
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4395058"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      190 a    253 c    239 g    222 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 172;  Length 904;
  Best Local Similarity   87.0%;  Pred. No. 5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     33 GAGCCTCGCCAGGTGCCGGTCAG 11


RESULT   9
AL580709/c
LOCUS       AL580709      917 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL580709 LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ008YM14 5
            prime, mRNA sequence.
ACCESSION   AL580709
VERSION     AL580709.1  GI:12946994
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 917)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .917
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YM14"
                     /clone_lib="LTI_NFL008_TC2"
                     /sex="male"
                     /tissue_type="T cells from T cell leukemia"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime  end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      204 a    246 c    233 g    229 t      5 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 106;  Length 917;
  Best Local Similarity   87.0%;  Pred. No. 5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     59 GAGCCTCGCCAGGTGCCGGTCAG 37


RESULT  10
BF979422/c
LOCUS       BF979422      1095 bp    mRNA            EST       23-JAN-2001
DEFINITION  602288251F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4373772 5',
            mRNA sequence.
ACCESSION   BF979422
VERSION     BF979422.1  GI:12346637
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10036  row: f  column: 13
            High quality sequence stop: 585.
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4373772"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      248 a    314 c    343 g    188 t      2 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 172;  Length 1095;
  Best Local Similarity   87.0%;  Pred. No. 5.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db    326 GAGCCTCGCCAGGTGCCGGTCAG 304


RESULT  11
AQ982558/c
LOCUS       AQ982558      494 bp    DNA             GSS       29-JAN-2000
DEFINITION  RPCI-23-294H19.TJ RPCI-23 Mus musculus genomic clone RPCI-23-294H19
            , DNA sequence.
ACCESSION   AQ982558
VERSION     AQ982558.1  GI:6814859
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-294H19.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 294  row: H  column: 19
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-294H19"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      167 a     94 c    130 g    102 t      1 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 236;  Length 494;
  Best Local Similarity   90.5%;  Pred. No. 7.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtca 23
          |||||||||||| ||||||  ||
Db     98 ATCCTCGTCAGTTGCAGGACA 78


RESULT  12
BE234470
LOCUS       BE234470      536 bp    mRNA            EST       10-JUL-2000
DEFINITION  141586 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE234470
VERSION     BE234470.1  GI:9019188
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 79  row: F  column: 18
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT       92 a    126 c    170 g    146 t      2 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 165;  Length 536;
  Best Local Similarity   90.5%;  Pred. No. 7.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtca 23
          ||||| | ||||||||||||||
Db    162 ATCCTGGGCAGGTGCAGGTCA 182


RESULT  13
```

```
AL538687
LOCUS       AL538687      777 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL538687 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF037YP11 3
            prime, mRNA sequence.
ACCESSION   AL538687
VERSION     AL538687.1  GI:12867208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DF037YP11"
                     /clone_lib="LTI_FL013_FBrn1"
                     /dev_stage="pooled tissue from post conception fetuses (20
                     week, 24 week and 26 week)"
                     /lab_host="DH10B"
                     /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector.  Library was constructed by Life
                     Technologies.  Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      173 a    220 c    205 g    169 t     10 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 106;  Length 777;
  Best Local Similarity   82.6%;  Pred. No. 7.5e+02;
  Matches   19;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          | |||||||||| : ||||||||
Db    486 GGTCCTCGTCAGAKCCAGGTCAG 508


RESULT  14
CNS044UQ
LOCUS       CNS044UQ     1032 bp    DNA             GSS       18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            082E16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL274571
VERSION     AL274571.1  GI:8008771
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1032)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="082E16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG082BC08LP1~end : T7"
BASE COUNT      267 a    249 c    220 g    261 t     35 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 221;  Length 1032;
  Best Local Similarity   90.5%;  Pred. No. 7.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcag 24
          |||||  ||||||||||||||
Db    753 TCCTCAGCAGGTGCAGGTCAG 773


RESULT  15
AA594461
LOCUS       AA594461      339 bp    mRNA            EST       25-SEP-1997
DEFINITION  nl94c07.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1058316 3'
            similar to gb:M77198 RAC-BETA SERINE/THREONINE KINASE (HUMAN);,
            mRNA sequence.
ACCESSION   AA594461
VERSION     AA594461.1  GI:2409811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 2573   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1058316"
```

```
                     /clone_lib="NCI_CGAP_Co10"
                     /tissue_type="colon tumor RER+"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     RER+ colon tumor, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo (N-Soares4). "
BASE COUNT       99 a     76 c     96 g     68 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 339;
  Best Local Similarity   83.3%;  Pred. No. 8.6e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           ||||| ||||| | ||| ||||||
Db      45 GATCCGCGTCACGCGCACGTCAGC 68


RESULT  16
BF318689/c
LOCUS       BF318689      360 bp    mRNA            EST       29-DEC-2000
DEFINITION  uy05a01.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone
            IMAGE:3657096 5' similar to TR:Q9VB28 Q9VB28 CG13972 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BF318689
VERSION     BF318689.1  GI:11267302
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1419400
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Mus musculus"
                     /strain="CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3657096"
                     /clone_lib="McCarrey Eddy spermatocytes"
                     /sex="male"
                     /tissue_type="spermatocytes, pooled from multiple mice"
                     /dev_stage="60 day"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
                     ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
                     [5'-(GA)10-ACTAGTCTCGAGTTTTTTTTTTTT-3'] and directionally
                     cloned  using 5' linkers 5'-AATTCGGCACGAG-3' and
                     5'-CTCGTGCCG-3'. Size selection of >400bp material gives
                     average insert size ranging from 1-2 kb. Library was mass
                     excised (from  lambda-UniZAP-XR) and resulting
                     single-stranded phagemids were prepped and tranformed
                     into DH10B. Library contains 98% recombinants.
                     References: J. Androl. 20:635-639 and Gene 25:263-269.
                     Library constructed and donated by J. McCarrey, Ph.D.
                     (Southwest Foundation for Biomedical Research, Dept. of
                     Genetics); excision done by E.M. Eddy, Ph.D. (National
                     Institutes of  Health, National Institute of Environmental
                     Health Sciences). Original lambda-based library is
                     available through ATCC, catalog #63422."
BASE COUNT      118 a     71 c     115 g     55 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 147;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 8.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
             |||| ||| |||| | |||||||
Db     181 CGATGCTCCTCAGCTNCAGGTCATC 157


RESULT  17
BF659032
LOCUS       BF659032      369 bp    mRNA            EST       20-DEC-2000
DEFINITION  maa89h11.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3824132 5' similar to SW:PI4K_HUMAN P42356
            PHOSPHATIDYLINOSITOL 4-KINASE ALPHA ;, mRNA sequence.
ACCESSION   BF659032
VERSION     BF659032.1  GI:11924166
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1460484
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 347.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3824132"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       70 a     121 c     104 g     74 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 151;  Length 369;
  Best Local Similarity   83.3%;  Pred. No. 8.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db     52 GAGCCTCGTCCGGTGCAGGACTGC 75


RESULT  18
BE860402
LOCUS       BE860402      398 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-AN1-afc-d-12-0-UI.r1 NIH_BMAP_MBG_N Mus musculus cDNA clone
            UI-M-AN1-afc-d-12-0-UI 5', mRNA sequence.
ACCESSION   BE860402
VERSION     BE860402.1  GI:10377301
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AN1-afc-d-12-0-UI"
                     /clone_lib="NIH_BMAP_MBG_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MBG_N library is a normalized library constructed
                     from  mouse basal ganglia. The tag is a string of 5
                     nucleotides present  between the Not I site and the
                     oligo-dT track.  The library was  constructed as described
                     by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                     , 1996. Tissue provided by Ms. Annie Novakovich,
                     Zivic-Miller Laboratories."
BASE COUNT       57 a    112 c    148 g     80 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 141;  Length 398;
  Best Local Similarity   83.3%;  Pred. No. 8.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db     78 GAGCCTCGTCCGGTGCAGGACTGC 101


RESULT  19
AA644735/c
LOCUS       AA644735      413 bp    mRNA            EST       28-OCT-1997
DEFINITION  M-EST008 Sugarcane mature stalk Saccharum sp. cDNA clone MA13-rev,
            mRNA sequence.
ACCESSION   AA644735
VERSION     AA644735.1  GI:2571144
KEYWORDS    EST.
SOURCE      Saccharum sp.
  ORGANISM  Saccharum sp.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Carson,D.L. and Botha,F.C.
  TITLE     Sugarcane cDNA from mature stalk tissue
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Deborah Lee Carson
            Biotechnology
            South African Sugar Association Experiment Station
            Private Bag XO2, Mount Edgecombe, 4300, South Africa
            Tel: 2731593205
            Fax: 2731595406
            Email: biotech@sugar.org.za
            Partial sequence of a 600 bp clone
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Saccharum sp."
                     /strain="variety Nco 376"
                     /db_xref="taxon:15819"
                     /clone="MA13-rev"
                     /clone_lib="Sugarcane mature stalk"
                     /note="Vector: Lambda ZAP II; Site_1: EcoRI; mRNA was
                     prepared from mature stalk material of the commercial
                     sugarcane variety Nco 376.  cDNA was synthesized using
                     standard protocols and ligated to an annealed
                     amplification adaptor set.   cDNA was then PCR amplified.
                     Amplified cDNA was ligated and cloned into the Eco RI site
                     of the Lambda ZAP II phage vector."
BASE COUNT      102 a    104 c     93 g    113 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 413;
  Best Local Similarity   83.3%;  Pred. No. 8.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          |||||| ||||| |||||||  ||
Db    188 GATCCTTGTCAGCTGCAGGTTTGC 165


RESULT  20
BE862745/c
LOCUS       BE862745      413 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-BH0-ajf-b-11-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-ajf-b-11-0-UI 5', mRNA sequence.
ACCESSION   BE862745
VERSION     BE862745.1  GI:10382015
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
```

```
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .413
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-BH0-ajf-b-11-0-UI"
                      /clone_lib="NIH_BMAP_M_S1"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_M_S1 library is a subtracted library derived from
                      a mixture of normalized libraries from ten regions of the
                      mouse brain (cerebellum, brain stems, olfactory bulbs,
                      hypothalamus, cortex, amygdala, basal ganglia, pineal
                      gland, striatum, hipoccampus). The driver used for
                      subtraction consisted of a pool of 20,000 cDNA clones
                      obtained from non-normalized and normalized libraries of
                      these ten regions of the mouse brain."
BASE COUNT       92 a    125 c    112 g     84 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 141;  Length 413;
  Best Local Similarity   83.3%;  Pred. No. 8.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | ||   ||||||||||||||||||
Db     24 CAATGTCCGTCAGGTGCAGGTCAG 1


RESULT  21
AQ172663/c
LOCUS       AQ172663      424 bp    DNA             GSS       17-OCT-1998
DEFINITION  HS_3194_A2_D09_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3194 Col=18 Row=G, DNA sequence.
ACCESSION   AQ172663
VERSION     AQ172663.1  GI:3570030
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3194  row: G  column: 18
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES              Location/Qualifiers
     source           1. .424
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3194 Col=18 Row=G"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      119 a    110 c     94 g    101 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 225;  Length 424;
  Best Local Similarity   83.3%;  Pred. No. 8.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | ||   ||||||||||||||||||
Db    382 CGATCTTGGCCAGGTGGAGGTCAG 359


RESULT  22
AQ371411/c
LOCUS       AQ371411      429 bp    DNA             GSS       06-MAR-1999
DEFINITION  HS_5043_A1_D05_SP6E RPCI11 Human Male BAC Library Homo sapiens
            genomic clone Plate=619 Col=9 Row=G, DNA sequence.
ACCESSION   AQ371411
VERSION     AQ371411.1  GI:4340516
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 619  row: G  column: 9
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 429.
FEATURES              Location/Qualifiers
     source           1. .429
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=619 Col=9 Row=G"
                      /clone_lib="RPCI11 Human Male BAC Library"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      122 a     90 c    109 g    107 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 227;  Length 429;
  Best Local Similarity   83.3%;  Pred. No. 8.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||| || || ||||||||||
Db     52 GAGCCTCGGCAAGTTCAGGTCAGC 29


RESULT  23
BE850719
LOCUS       BE850719      463 bp    mRNA            EST       26-SEP-2000
DEFINITION  ux01a08.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3470198 5', mRNA sequence.
ACCESSION   BE850719
VERSION     BE850719.1  GI:10309058
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1389558
            Seq primer: -40RP from Gibco
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3470198"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       85 a    123 c    173 g     82 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 141;  Length 463;
  Best Local Similarity   83.3%;  Pred. No. 8.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    233 GAGCCTCGTCCGGTGCAGGACTGC 256


RESULT  24
AI789420/c
LOCUS       AI789420      495 bp    mRNA            EST       02-JUL-1999
DEFINITION  uk99a06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2064658 5', mRNA sequence.
ACCESSION   AI789420
VERSION     AI789420.1  GI:5337136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uk99a06.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:993846
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2064658"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      147 a    122 c    141 g     85 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 24;  Length 495;
  Best Local Similarity   83.3%;  Pred. No. 8.9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | ||   ||||||||||||||||||
Db    400 CAATGTCCGTCAGGTGCAGGTCAG 377


RESULT  25
AI180705
LOCUS       AI180705      508 bp    mRNA            EST       08-OCT-1998
DEFINITION  ub91b01.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1395817 5' similar to TR:Q13288 Q13288 P37NB. ;, mRNA
            sequence.
ACCESSION   AI180705
VERSION     AI180705.1  GI:3731343
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
```

```
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:907533
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1395817"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      112 a    135 c    146 g    114 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 17;  Length 508;
  Best Local Similarity   83.3%;  Pred. No. 8.9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          ||||||||| | | |||||| |||
Db     39 CGATCCTCGCCTGCTGCAGGACAG 62


RESULT  26
AA536738
LOCUS       AA536738      618 bp    mRNA            EST       29-JUL-1997
DEFINITION  vj88c01.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            IMAGE:944160 3' similar to SW:PI4K_HUMAN P42356
            PHOSPHATIDYLINOSITOL 4-KINASE ALPHA ;, mRNA sequence.
ACCESSION   AA536738
VERSION     AA536738.1  GI:2282731
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:541016
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:944160"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos. Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'. cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe vector using commercial linkers (NEB).
                     Average insert size: 1.2 kb."
BASE COUNT      149 a    148 c    224 g     95 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 8;  Length 618;
  Best Local Similarity   83.3%;  Pred. No. 9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    261 GAGCCTCGTCCGGTGCAGGACTGC 284


RESULT  27
BE653574
LOCUS       BE653574      628 bp    mRNA            EST       06-SEP-2000
DEFINITION  UI-M-AM1-agb-h-08-0-UI.r1 NIH_BMAP_MAM_N Mus musculus cDNA clone
            UI-M-AM1-agb-h-08-0-UI 5', mRNA sequence.
ACCESSION   BE653574
VERSION     BE653574.1  GI:9979482
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
```

```
           Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AM1-agb-h-08-0-UI"
                     /clone_lib="NIH_BMAP_MAM_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MAM_N library is a normalized library constructed
                     from  mouse amygdala. The tag is a string of 5 nucleotides
                     present  between the Not I site and the oligo-dT track.
                     The library was  constructed as described by Bonaldo,
                     Lennon and Soares,  Genome Research 6: 791-806, 1996.
                     Tissue provided by Ms. Annie Novakovich, Zivic-Miller
                     Laboratories."
BASE COUNT      121 a    172 c    207 g    128 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 138;  Length 628;
  Best Local Similarity   83.3%;  Pred. No. 9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    430 GAGCCTCGTCCGGTGCAGGACTGC 453


RESULT  28
AZ426129/c
LOCUS       AZ426129      641 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0206H03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0206H03 R, DNA sequence.
ACCESSION   AZ426129
VERSION     AZ426129.1  GI:10550142
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0206  row: H  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0206H03"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      156 a    184 c    183 g    118 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 243;  Length 641;
  Best Local Similarity   83.3%;  Pred. No. 9.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | ||   ||||||||||||||||
Db    219 CAATGTCCGTCAGGTGCAGGTCAG 196


RESULT  29
BE262905/c
LOCUS       BE262905      721 bp    mRNA            EST       26-OCT-2000
DEFINITION  601153594F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3509542 5',
            mRNA sequence.
ACCESSION   BE262905
VERSION     BE262905.1  GI:9136326
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM192  row: l  column: 23
            High quality sequence start: 56
            High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3509542"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
```

California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 186 a 216 c 200 g 119 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 165; Length 721;
Best Local Similarity 83.3%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
| ||||||| ||||||||| |||
Db 338 GGTCCTCGTGTGGTGCAGGTAAGC 315

RESULT 30
BG478947/c
LOCUS BG478947 833 bp mRNA EST 21-MAR-2001
DEFINITION 602525709F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649354 5',
mRNA sequence.
ACCESSION BG478947
VERSION BG478947.1 GI:13411226
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1429 row: a column: 03
High quality sequence stop: 622.
FEATURES Location/Qualifiers
source 1. .833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4649354"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 178 a 291 c 263 g 101 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 154; Length 833;
Best Local Similarity 83.3%; Pred. No. 9.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcag 24
|| |||||||| ||||| |||| |
Db 572 CGGTCCTCGTCTGGTGCTGGTCCG 549

RESULT 31
BF794272/c
LOCUS BF794272 893 bp mRNA EST 12-JAN-2001
DEFINITION 602255473F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4338715 5',
mRNA sequence.
ACCESSION BF794272
VERSION BF794272.1 GI:12099326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 893)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9948 row: a column: 20
High quality sequence stop: 672.
FEATURES Location/Qualifiers
source 1. .893
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4338715"
/clone_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 204 a 229 c 264 g 196 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 169; Length 893;
Best Local Similarity 83.3%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
| ||||| ||||| ||||| ||||
Db 584 GCTCCTCCTCAGGAGCAGGGCAGC 561

RESULT 32
CNS02FRJ
LOCUS CNS02FRJ 897 bp DNA GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
135001 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL195400
VERSION AL195400.1 GI:7833506
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 897)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 897)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="135O01"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG135AH01LP1~end : T7"
BASE COUNT      185 a    227 c    259 g    223 t      3 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 220;  Length 897;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||||| ||||| ||||| |||
Db    808 GATCCTCTGCAGGTTCAGGTGAGC 831


RESULT  33
BF313136/c
LOCUS       BF313136      929 bp    mRNA            EST       21-NOV-2000
DEFINITION  601897654F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127005 5',
            mRNA sequence.
ACCESSION   BF313136
VERSION     BF313136.1  GI:11261079
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1017  row: l  column: 14
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4127005"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      232 a    214 c    280 g    203 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 147;  Length 929;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          | ||||| ||||| ||||| ||||
Db    599 GCTCCTCCTCAGGAGCAGGGCAGC 576


RESULT  34
CNS00L1P
LOCUS       CNS00L1P      942 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR23G17 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL067857
VERSION     AL067857.1  GI:4957886
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR23G17"
                     /note="end : T7"
BASE COUNT      292 a    166 c    183 g    242 t     59 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 219;  Length 942;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||| || |||| |||||| |||||
Db    191 GATTCTGGTCATGTGCAGATCAGC 214
```

```
RESULT  35
CNS006PE/c
LOCUS       CNS006PE      982 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065656
VERSION     AL065656.1  GI:4944725
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14J05"
                     /note="end : T7"
BASE COUNT      205 a    289 c    287 g    161 t     40 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 219;  Length 982;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||| |||| || || |||||||
Db    116 GATCCACGTCGGGAGCGGGTCAGC 93


RESULT  36
CNS00HWP/c
LOCUS       CNS00HWP      985 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL073836
VERSION     AL073836.1  GI:4953810
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 985)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .985
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR35J21"
                     /note="end : TET3"
BASE COUNT      204 a    275 c    300 g    164 t     42 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 219;  Length 985;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||| |||| || || |||||||
Db    117 GATCCACGTCGGGAGCGGGTCAGC 94


RESULT  37
BF965985/c
LOCUS       BF965985      1002 bp    mRNA            EST       22-JAN-2001
DEFINITION  602277168F2 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4364968 5',
            mRNA sequence.
ACCESSION   BF965985
VERSION     BF965985.1  GI:12333200
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10013  row: g  column: 17
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .1002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4364968"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
```

```
                    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                    Average insert size 1.533 kb. Library enriched for
                    full-length clones and constructed by Life Technologies.
                    Note: this is a NIH_MGC Library."
BASE COUNT      266 a    266 c    288 g    182 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 171;  Length 1002;
  Best Local Similarity   83.3%;  Pred. No. 9.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          | ||||| ||||| ||||| ||||
Db    649 GCTCCTCCTCAGGAGCAGGGCAGC 626


RESULT  38
CNS00HQI/c
LOCUS       CNS00HQI     1016 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35L24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL074118
VERSION     AL074118.1  GI:4953595
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1016
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR35L24"
                      /note="end : TET3"
BASE COUNT      201 a    295 c    299 g    169 t     52 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 219;  Length 1016;
  Best Local Similarity   83.3%;  Pred. No. 9.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||| |||| || || |||||||
Db    118 GATCCACGTCGGGAGCGGGTCAGC 95


RESULT  39
```

```
BE795069/c
LOCUS       BE795069     1126 bp    mRNA            EST       20-SEP-2000
DEFINITION  601592885F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946869 5',
            mRNA sequence.
ACCESSION   BE795069
VERSION     BE795069.1  GI:10216267
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1126)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM807  row: j  column: 22
            High quality sequence start: 10
            High quality sequence stop: 702.
FEATURES              Location/Qualifiers
     source           1. .1126
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3946869"
                      /clone_lib="NIH_MGC_7"
                      /tissue_type="small cell carcinoma"
                      /cell_line="MGC3"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      262 a    320 c    289 g    255 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 140;  Length 1126;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcagg 20
          |||||||||||||||| |||
Db    853 GATCCTCGTCAGGTGTAGG 835


RESULT  40
BG328084/c
LOCUS       BG328084     1298 bp    mRNA            EST       27-FEB-2001
DEFINITION  602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
            mRNA sequence.
ACCESSION   BG328084
VERSION     BG328084.1  GI:13134522
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1298)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1233  row: c  column: 10
            High quality sequence stop: 511.
FEATURES              Location/Qualifiers
     source           1. .1298
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4546497"
                      /clone_lib="NIH_MGC_15"
                      /tissue_type="adenocarcinoma cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      316 a    498 c    357 g    127 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 152;  Length 1298;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtc 22
          || |||||||||||||||||
Db    987 TCGTCGTCAGGTGCAGGTC 969


RESULT  41
AW371554
LOCUS       AW371554      108 bp    mRNA            EST       04-FEB-2000
DEFINITION  QV3-BT0295-111199-017-g08 BT0295 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW371554
VERSION     AW371554.1  GI:6876208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-BT0295-
            111199-017-g08&t3=1999-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 108.
FEATURES              Location/Qualifiers
     source           1. .108
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0295"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT        8 a     39 c     23 g     38 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 115;  Length 108;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db      1 TCCTCCTCAGGAGCAGGGCAGC 22


RESULT  42
AA651689/c
LOCUS       AA651689      112 bp    mRNA            EST       13-NOV-1997
DEFINITION  nn47b06.r1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1086995 5'
            similar to gb:X63526 ELONGATION FACTOR 1-GAMMA (HUMAN);, mRNA
            sequence.
ACCESSION   AA651689
VERSION     AA651689.1  GI:2583341
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 815   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .112
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1086995"
                      /clone_lib="NCI_CGAP_Kid6"
                      /sex="mixed"
                      /tissue_type="kidney tumor"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size:  1.0 kb."
BASE COUNT       44 a     21 c     36 g     11 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 112;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db    101 TCCTCCTCAGGAGCAGGGCAGC 80


RESULT  43
AA197041/c
LOCUS       AA197041      235 bp    mRNA            EST       12-MAR-1998
DEFINITION  zq10c11.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
            IMAGE:629300 5' similar to gb:X63526 ELONGATION FACTOR 1-GAMMA
            (HUMAN);, mRNA sequence.
ACCESSION   AA197041
VERSION     AA197041.1  GI:1792665
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 235)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 757   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 138.
FEATURES             Location/Qualifiers
     source          1. .235
                     /organism="Homo sapiens"
                     /db_xref="GDB:5049534"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:629300"
                     /clone_lib="Stratagene muscle 937209"
                     /tissue_type="muscle"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skeletal muscle; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
                     Primer: Oligo dT. Skeletal muscle from patient with
                     malignant hyperthermia.  Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       56 a     52 c     71 g     45 t     11 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 3;  Length 235;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db     83 TCCTCCTCAGGAGCAGGGCAGC 62


RESULT  44
BF349456
LOCUS       BF349456      236 bp    mRNA            EST       22-NOV-2000
DEFINITION  RC2-HT0148-021099-012-g03 HT0148 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF349456
VERSION     BF349456.1  GI:11308530
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC2&t2=RC2-HT0148-
            021099-012-g03&t3=1999-10-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 8.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0148"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     40 c     58 g     70 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 147;  Length 236;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          |||||||||||||  ||||| |
Db     47 TCCTCGTCAGGTGATGGTCATC 68


RESULT  45
N83353/c
LOCUS       N83353        255 bp    mRNA            EST       01-APR-1996
DEFINITION  EC0312F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
            clone EC0312 5' similar to ELONGATION FACTOR 1-GAMMA, mRNA
            sequence.
ACCESSION   N83353
VERSION     N83353.1  GI:1258978
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 255)
  AUTHORS   Liew,C.C.
  TITLE     cDNAs from fetal heart (1996)
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Liew CC
            Department of Laboratory Medicine and Pathobiology
```

University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Seq primer: GAAATTAACCCTCACTAAAGGG.

```
FEATURES             Location/Qualifiers
     source          1. .255
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="EC0312"
                     /clone_lib="Human fetal heart, Lambda ZAP Express"
                     /lab_host="E. coli XL1-Blue"
                     /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was purified from human fetal hearts (8-10
                     weeks). cDNA was synthesized using a XhoI-Oligo dT
                     adaptor-primer. EcoRI adaptors were ligated, followed by
                     digestion with XhoI, for directional cloning into
                     predigested lambda ZAP Express."
BASE COUNT       76 a     58 c     80 g     41 t
ORIGIN
```

Query Match 68.8%; Score 17.2; DB 159; Length 255;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       4 tcctcgtcaggtgcaggtcagc 25
           ||||| ||||| ||||| ||||
Db     229 TCCTCCTCAGGAGCAGGGCAGC 208
```

Search completed: October 9, 2001, 13:46:52
Job time: 9527 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:14 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 188662

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 BioB gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

	9	25	100.0	5872	9	A93679 A93679 Sequence 6
	10	25	100.0	5872	9	AR101809 AR101809 Sequence
	11	25	100.0	5872	9	AR101810 AR101810 Sequence
	12	25	100.0	11022	1	AE000180 AE000180 Escherich
	13	25	100.0	13501	1	AE005258 AE005258 Escherich
	14	25	100.0	297816	2	AP002553 AP002553 Escherich
	15	21	84.0	5526	2	AF250776 AF250776 Unculture
	16	18	72.0	8227	2	AF248314 AF248314 Unculture
c	17	16	64.0	32767	77	AC087305 AC087305 Homo sapi
	18	16	64.0	155172	66	AC021110 AC021110 Homo sapi
c	19	16	64.0	156184	71	AC041018 AC041018 Homo sapi
c	20	16	64.0	174815	82	AP001390 AP001390 Homo sapi
c	21	16	64.0	196172	66	AC020734 AC020734 Homo sapi
	22	15	60.0	1771	9	AR080743 AR080743 Sequence
	23	15	60.0	1771	94	MUSINTESUB L20788 Mus musculu
c	24	15	60.0	1829	54	G07181 G07181 human STS W
	25	15	60.0	1949	94	MMINTEGR01 U34627 Mus musculu
	26	15	60.0	2340	2	ECHK31IM X82231 E.coli HK31
c	27	15	60.0	3748	10	AX098224 AX098224 Sequence
c	28	15	60.0	4602	91	D13644 D13644 Human mRNA
	29	15	60.0	4636	1	AF039582 AF039582 Enterobac
c	30	15	60.0	5218	89	AF152313 AF152313 Homo sapi
c	31	15	60.0	5221	89	AF152315 AF152315 Homo sapi
c	32	15	60.0	5233	89	AF152308 AF152308 Homo sapi
c	33	15	60.0	5251	89	AF152312 AF152312 Homo sapi
c	34	15	60.0	5254	89	AF152306 AF152306 Homo sapi
c	35	15	60.0	5254	89	AF152310 AF152310 Homo sapi
c	36	15	60.0	5257	89	AF152307 AF152307 Homo sapi
c	37	15	60.0	5260	89	AF152305 AF152305 Homo sapi
c	38	15	60.0	5260	89	AF152309 AF152309 Homo sapi
c	39	15	60.0	5260	89	AF152311 AF152311 Homo sapi
c	40	15	60.0	5260	89	AF152314 AF152314 Homo sapi
c	41	15	60.0	5260	89	AF152316 AF152316 Homo sapi
c	42	15	60.0	5260	89	AF152317 AF152317 Homo sapi
c	43	15	60.0	5299	89	AF152303 AF152303 Homo sapi
c	44	15	60.0	5431	89	AF152304 AF152304 Homo sapi
	45	15	60.0	5562	85	AB011138 AB011138 Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   2
```

```
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT       271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
```

```
Db     44 GACATTGTCGCAAGTCACAGAATTA 68


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 10;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      62 GACATTGTCGCAAGTCACAGAATTA 86


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 2;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db   2032 GACATTGTCGCAAGTCACAGAATTA 2056


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
```

```
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db    137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db    137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"

BASE COUNT     1318 a,  1552 c,  1695 g   1307 t

ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
```

YINLAFVDASTLGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
||||||||||||||||||||||||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 10
AR101809
LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
||||||||||||||||||||||||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 11
AR101810
LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
||||||||||||||||||||||||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 12
AE000180
LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"

```
                     /db_xref="taxon:83333"
repeat_region        <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
gene                 complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
CDS                  complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
                     /protein_id="AAC73859.1"
                     /db_xref="GI:1786989"
                     /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                     LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                     AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                     DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                     LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                     NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                     AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                     DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter             complement(1507. .1535)
                     /note="factor Sigma70; predicted +1 start at 806574"
gene                 complement(1582. .2058)
                     /gene="ybhB"
                     /note="b0773"
CDS                  complement(1582. .2058)
                     /gene="ybhB"
                     /function="orf; Unknown"
                     /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC73860.1"
                     /db_xref="GI:1786990"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind         complement(2108. .2124)
                     /note="central position to predicted promoter:85.5"
                     /bound_moiety="RhaS predicted site"
gene                 complement(2117. .3406)
                     /gene="bioA"
                     /note="b0774"
CDS                  complement(2117. .3406)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diaminopelargonic acid synthetase"
                     /protein_id="AAC73861.1"
                     /db_xref="GI:1786991"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter             complement(2193. .2221)
                     /gene="bioA"
                     /note="factor Sigma70; predicted +1 start at 807260"
protein_bind         3411. .3450
                     /note="central position to bioB promoter: -20"
                     /bound_moiety="BioB documented site"
protein_bind         3411. .3450
                     /note="central position to predicted promoter: -20;genetic
                     evidence for the site"
                     /bound_moiety="BioB predicted site"
protein_bind         complement(3411. .3450)
                     /note="central position to predicted promoter:50"
                     /bound_moiety="BioB predicted site"
protein_bind         complement(3411. .3450)
                     /note="central position to bioA promoter:50"
                     /bound_moiety="BioB documented site"
promoter             3413. .3441
                     /note="factor Sigma70; promoter bioB; documented +1 at
                     808525"
promoter             complement(3447. .3473)
                     /note="factor Sigma70; promoter bioA; documented +1 at
                     808515"
gene                 3493. .4533
                     /gene="bioB"
                     /note="b0775"
CDS                  3493. .4533
                     /gene="bioB"
                     /EC_number="2.8.1.-"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis, sulfur insertion?"
                     /protein_id="AAC73862.1"
                     /db_xref="GI:1786992"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
gene                 4530. .5684
                     /gene="bioF"
                     /note="b0776"
CDS                  4530. .5684
                     /gene="bioF"
                     /EC_number="2.3.1.47"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAC73863.1"
                     /db_xref="GI:1786993"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL

  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-05;
  Matches    25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db    3513 GACATTGTCGCAAGTCACAGAATTA 3537


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
```

```
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
```

```
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match            100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db    9599 GACATTGTCGCAAGTCACAGAATTA 9623


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites),
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
```

```
gene            1245. .2114
                /gene="ECs0754"
CDS             1245. .2114
                /gene="ECs0754"
                /note="similar to SUCD_ECOLI gi|1786949 percent identity
                100 in 289 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="succinyl-CoA synthetase alpha subunit"
                /protein_id="BAB34177.1"
                /db_xref="GI:13360213"
                /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
gene            2781. .3686
                /gene="ECs0755"
CDS             2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

```
  Query Match             100.0%;  Score 25;  DB 2;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db   92331 GACATTGTCGCAAGTCACAGAATTA 92355


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
```

```
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 2;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.0045;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcacagaatta 25
           |||||||||||||||||||||
Db    1987 TTGTCGCAAGTCACAGAATTA 2007


RESULT  16
AF248314
LOCUS       AF248314     8227 bp    DNA             BCT       24-JAN-2001
DEFINITION  Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
            partial cds; histidine ammonia-lyase-like protein (hutH), DAPA
            aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
            synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
            and dethiobiotin synthase BioD (bioD) genes, complete cds; ABC
            transporter-like protein (elsA) gene, partial cds; and unknown
            gene.
ACCESSION   AF248314
VERSION     AF248314.1  GI:12407610
KEYWORDS    .
SOURCE      uncultured bacterium pCosAS1.
  ORGANISM  uncultured bacterium pCosAS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct Cloning from Enrichment Cultures, a Reliable Strategy for
            Isolation of Complete Operons and Genes from Microbial Consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
   PUBMED   11133432
REFERENCE   2  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
source          1. .8227
                /organism="uncultured bacterium pCosAS1"
                /db_xref="taxon:143793"
                /clone="pCosAS1"
                /note="unknown organism, cosmid clone derived from
                environmental consortium"
gene            <1. .502
                /gene="hutU"
CDS             <1. .502
                /gene="hutU"
                /note="putative urocanate hydratase"
                /codon_start=2
                /transl_table=11
                /product="urocanase-like protein"
                /protein_id="AAG53593.1"
                /db_xref="GI:12407618"
                /translation="PRERISFQGLPARICWVGLEWRQKLGLAFNEMVRRGEVSAPIVI
                GRDHLDSGSVASPNRETEAMQDGSDAVSDWPLLNALLNTASGATWVSLHHGGGVGMGF
                SQHAGMVIVCDGTDDAAARIARVLHNDPATGVMRHADAGYDIAIDCAKEQGLNLPMIA
                ATQGQH"
gene            502. .2031
                /gene="hutH"
CDS             502. .2031
                /gene="hutH"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="histidine ammonia-lyase-like protein"
                /protein_id="AAG53586.1"
                /db_xref="GI:12407611"
                /translation="MNALTLTPGTLTLAQLRQVWQQPLQLTLDESAHEAINDSVACVE
                AIVAEGRTAYGINTGFGLLAQTRIATHDLENLQRSLVLSHAAGVGEPLDDDIVRLMMV
                LKINSLARGFSGIRLSVIQALIALVNAGVYSVDPAKGSVGASGDLAPLAHMSLTLLGE
                GKARYRGEWLPAATALQKAGLAPVTLAAKEGLALLNGTQASTAFALRGLFEAEDLFAS
                AVVCGALTTEAVLGSRRPFDARIHEVRGQRGQIDAAALFRHVLTDTSAIAESHHNCDK
                VQDPYSLRCQPQVMGACLTQMRQVAEVLLVESNAVSDNPLVFAAENEMVFRGNFHAEP
                VAMAADNLALAIAEIGALSERRIALMMDKHMSQLPPFLVRNGGVNSGFMIAQVTAAAL
                ASENKGLCHPHTTSVDKPCRPSANQEDHVSMAPAAGRRLWEMAGNTRGVLAVEWLAAC
                QGADLRDGLTSSPLLEQARQSCGEQVAHYDDDRFFAPDIEAAISLLNKGSLVGLLPAF
                L"
CDS             complement(2130. .2606)
                /note="ORF1; similar to hypothetical 17.1 kDa protein in
                modC-bioA intergenic region"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAG53587.1"
                /db_xref="GI:12407612"
                /translation="MKLISQDLRDGDKLPLRHVFNGMGYEGDNISPHLAWDEVPSGTK
                SFVVTCYDPDAPTDSGWWHWIVANIPADTRVLPQGSGSSLVALPEGAVETRTDFGKAG
                YGGAAPPKGETHRYIFTVHALNVERIEVDEEASGAMVGFNVHFHTLGSASITAMYS"
gene            complement(2651. .3940)
                /gene="bioA"
CDS             complement(2651. .3940)
                /gene="bioA"
                /note="7,8-diamino-pelargonic acid aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="DAPA aminotransferase BioA"
                /protein_id="AAG53588.1"
                /db_xref="GI:12407613"
                /translation="MTTDDLAFDRQHIWHPYTSMTQPLPVYPVVEAHGCELTLASGEQ
                LVDGMSSWWAAIHGYNHPRLNAAMKAQIDRMSHVMFGGITHQPAVDLCRRLVAMTPET
                LECVFPADSGSVAVEVAMKMALQYWHAKGESRQRFITFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRFDGAWDEMDMVGLAPLMAAHRHEIAAVILEPIVQGAG
                GMRMYHPEWLKRIRRMCDREGILLIADEIATGFGRTGKLFACEHAGITPDILCLGKAL
                TGGTMTMSATLTTRQVAETISNGEAGCFMHGPTFMGNPLACAVASESLALLESGEWQD
                QVAAIESQLQEELAVARDSEFVADVRVLGAIGVVETTHPVNMAALQRFFVARGVWVRP
                FGKLIYLMPPYVIRPDQLSKLTRAVVDAVQAPAHFTR"
gene            4029. .5069
                /gene="bioB"
CDS             4029. .5069
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase BioB"
                /protein_id="AAG53589.1"
                /db_xref="GI:12407614"
                /translation="MAHQARWTLSQVTALFEKPLLELLFEAQQIHRQHFDPQQIQVST
                LLSIKTGACPEDCKYCPQSARYKTGLESERLMEVEQVLESARQAKNAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLQACMTLGTLDDHQAQRLASAGLDYYNHNLDTFPEF
                YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLCETVTDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVXAFDXIRTXSVAXVMXPTXLVRLSTGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTAVLTGDNEQQHQLEQQLFN
                ADTDQFYNAATV"
gene            5066. .6223
                /gene="bioF"
CDS             5066. .6223
                /gene="bioF"
                /note="7-keto-8-amino-pelargonic acid synthetase; similar
                to BioF"
                /codon_start=1
                /transl_table=11
                /product="7-KAPA synthetase"
                /protein_id="AAG53590.1"
                /db_xref="GI:12407615"
                /translation="MTWQQRIDAALDARRAADALRVRTPVENGAGRFLIREQRRYCNF
                SGNDYLGLSQHPEIIRAWQQGAEQYGVGSGGSGHVSGYTHAHQALEEHLAGWLGYPRA
                LLFISGFAANQALIAALTDKEDRIVADRLSHASLLEAASHSPAQLRRFAHNDVAQLDS
                LLDKPCHGQQLVVTEGIFSMDGDSAPLGVIHAAAQRQGSWLLVDDAHGIGVTGPEGRG
                SAASQGVKPELLVVTFGKGVGVSGAAVLCSDAVADYLLQFARHLIYSTSMPPAQAVAL
                LAAFNVIRSDEDNARRQRLAALIDQFRAGVRDLPVQITASQSAIQPLIVGENARALHL
                AQRLRERGIWVTAIRPPTVPAGTARLRLTLTAAHQPADITALVEALHVAGE"
gene            6207. .6962
                /gene="bioC"
CDS             6207. .6962
                /gene="bioC"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis BioC-like protein"
                /protein_id="AAG53591.1"
                /db_xref="GI:12407616"
                /translation="MSQVNKQAIAAAFGRAAQSYSRHDALQRQSADGLRALLAENTFA
                SVLDAGCGPGSNSLAWRQTGSVVTAMDLCAPMLEEARRNQAADSYLQADIEAIPLADA
                QFSLAWSHLAVQXCASLPRALAELWRVTQLGGKVAFTTLLNGSLPELNQAWQAVDTQP
                HANRFLSGQQVNDALSGWQYRCAVQTVTLLFDDALSAMHSLKGIGATHLHAGRAAKPL
                TRGQLQQLAQAWPQQAGKFPLSYQLFHGIIERD"
gene            6940. .7680
                /gene="bioD"
CDS             6940. .7680
                /gene="bioD"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthase BioD"
                /protein_id="AAG53592.1"
                /db_xref="GI:12407617"
                /translation="MELSNVTERYFVTGTDTEVGKTVASTALLQAARALGRHTAGYKP
                VASGSEMTAQGLRNSDALALQRNSSLALSYPAVNPYTFAEPTSPHIVSADEQRPIEFA
                TLSAGLRALETQADWVLVEGAGGWFTPLSQEQTFADWVMAERLPVILVVGVKLGCINH
                AMLTAQAIRQAGLHFAGWVANGVVVPGKRHAEYMATLRRVLSAPLLGEIPWLGEQADT
                SAVGQYLDLTALTPSAPSSGQSAGHPAR"
gene            complement(7619. .>8227)
                /gene="elsA"
CDS             complement(7619. .>8227)
                /gene="elsA"
                /note="putative urea/short-chain amide ABC transporter"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter-like protein"
                /protein_id="AAG53594.1"
                /db_xref="GI:12407619"
                /translation="DCXLRTRXGENXLVSCIAGLLPIASGTMVWHEAPGAVNQAGTEI
                XVVPQDRRIFSQLTVEENLHIAQMAKGIAGGSLSRDVFEFFPELYPLRQLRASGLNDD
                QQYQLALANALVTRPRLLILDEPSRGTGQLFTVKLAQLLLRFNRELGLSVLLAEQHLS
                LIRRVADRFCLLYRGRNVAQGRVSELDDPLIAHWMGPTASAQ"
BASE COUNT      1512 a   2642 c   2605 g   1453 t     15 others
```

```
ORIGIN


  Query Match             72.0%;  Score 18;  DB 2;  Length 8227;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcac 18
           ||||||||||||||||||
Db    4049 GACATTGTCGCAAGTCAC 4066


RESULT  17
AC087305/c
LOCUS       AC087305    32767 bp    DNA             HTG       23-DEC-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC087305
VERSION     AC087305.1  GI:11990696
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-238P13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12071
                Center clone name: 238_P_13
            --------------.
            * NOTE: This record contains 40 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However,,it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      707: contig of 707 bp in length
            *      708 807: gap of      100 bp
            *      808     1527: contig of 720 bp in length
            *     1528 1627: gap of      100 bp
            *     1628     2348: contig of 721 bp in length
            *     2349 2448: gap of      100 bp
            *     2449     3157: contig of 709 bp in length
            *     3158 3257: gap of      100 bp
            *     3258     3973: contig of 716 bp in length
            *     3974 4073: gap of      100 bp
            *     4074     4799: contig of 726 bp in length
            *     4800 4899: gap of      100 bp
            *     4900     5617: contig of 718 bp in length
            *     5618 5717: gap of      100 bp
            *     5718     6450: contig of 733 bp in length
            *     6451 6550: gap of      100 bp
            *     6551     7259: contig of 709 bp in length
            *     7260 7359: gap of      100 bp
            *     7360     8066: contig of 707 bp in length
            *     8067 8166: gap of      100 bp
            *     8167     8907: contig of 741 bp in length
            *     8908 9007: gap of      100 bp
            *     9008     9716: contig of 709 bp in length
            *     9717 9816: gap of      100 bp
            *     9817    10532: contig of 716 bp in length
            *    10533 10632: gap of      100 bp
            *    10633    11354: contig of 722 bp in length
            *    11355 11454: gap of      100 bp
            *    11455    12166: contig of 712 bp in length
            *    12167 12266: gap of      100 bp
            *    12267    12976: contig of 710 bp in length
            *    12977 13076: gap of      100 bp
            *    13077    13789: contig of 713 bp in length
            *    13790 13889: gap of      100 bp
            *    13890    14617: contig of 728 bp in length
            *    14618 14717: gap of      100 bp
            *    14718    15452: contig of 735 bp in length
            *    15453 15552: gap of      100 bp
            *    15553    16276: contig of 724 bp in length
            *    16277 16376: gap of      100 bp
            *    16377    17111: contig of 735 bp in length
            *    17112 17211: gap of      100 bp
            *    17212    17925: contig of 714 bp in length
            *    17926 18025: gap of      100 bp
            *    18026    18746: contig of 721 bp in length
            *    18747 18846: gap of      100 bp
            *    18847    19569: contig of 723 bp in length
            *    19570 19669: gap of      100 bp
            *    19670    20384: contig of 715 bp in length
            *    20385 20484: gap of      100 bp
            *    20485    21190: contig of 706 bp in length
            *    21191 21290: gap of      100 bp
            *    21291    22028: contig of 738 bp in length
            *    22029 22128: gap of      100 bp
            *    22129    22841: contig of 713 bp in length
            *    22842 22941: gap of      100 bp
            *    22942    23653: contig of 712 bp in length
            *    23654 23753: gap of      100 bp
            *    23754    24478: contig of 725 bp in length
            *    24479 24578: gap of      100 bp
            *    24579    25305: contig of 727 bp in length
            *    25306 25405: gap of      100 bp
            *    25406    26143: contig of 738 bp in length
            *    26144 26243: gap of      100 bp
            *    26244    26968: contig of 725 bp in length
            *    26969 27068: gap of      100 bp
            *    27069    27785: contig of 717 bp in length
            *    27786 27885: gap of      100 bp
            *    27886    28584: contig of 699 bp in length
            *    28585 28684: gap of      100 bp
```

```
            *    28685    29467: contig of 783 bp in length
            *    29468 29567: gap of      100 bp
            *    29568    30292: contig of 725 bp in length
            *    30293 30392: gap of      100 bp
            *    30393    31117: contig of 725 bp in length
            *    31118 31217: gap of      100 bp
            *    31218    31944: contig of 727 bp in length
            *    31945 32044: gap of      100 bp
            *    32045    32767: contig of 723 bp in length.
FEATURES             Location/Qualifiers
     source          1. .32767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-238P13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     8045 a    5670 c    5396 g    9330 t    4326 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 77;  Length 32767;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   30730 GCAAGTCACAGAATTA 30715


RESULT  18
AC021110
LOCUS       AC021110   155172 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC021110
VERSION     AC021110.3  GI:8439978
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 10, 2000 this sequence version replaced gi:7021826.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0238P13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 73%
            Sequencing vector: plasmid; 27%
            Chemistry: Dye-primer ET; 73% of reads
            Chemistry: Dye-terminator Big Dye; 27% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151162 bases at least Q40
            Consensus quality: 152443 bases at least Q30
            Consensus quality: 153249 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 154672; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.69 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9265: contig of 9265 bp in length
            *     9266     9365: gap of unknown length
            *     9366    26504: contig of 17139 bp in length
            *    26505    26604: gap of unknown length
            *    26605    44300: contig of 17696 bp in length
            *    44301    44400: gap of unknown length
            *    44401    62250: contig of 17850 bp in length
            *    62251    62350: gap of unknown length
            *    62351   105744: contig of 43394 bp in length
            *   105745   105844: gap of unknown length
            *   105845   155172: contig of 49328 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-238P13"
     misc_feature    1. .9265
                     /note="assembly_name:Contig1"
     misc_feature    9366. .26504
                     /note="assembly_name:Contig2"
     misc_feature    26605. .44300
                     /note="assembly_name:Contig3
                     clone_end:T7
                     vector_side:right"
     misc_feature    44401. .62250
                     /note="assembly_name:Contig4"
     misc_feature    62351. .105744
                     /note="assembly_name:Contig5"
     misc_feature    105845. .155172
                     /note="assembly_name:Contig6"
BASE COUNT     51543 a   28385 c   28043 g   46698 t    503 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 66;  Length 155172;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db  111929 GCAAGTCACAGAATTA 111944


RESULT  19
AC041018/c
LOCUS       AC041018   156184 bp    DNA             HTG       22-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-478O20 map 4, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC041018
VERSION     AC041018.2  GI:8016453
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-478O20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
```

```
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7534220.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9577
                Center clone name: 478_O_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147958 bases at least Q40
                Consensus quality: 151874 bases at least Q30
                Consensus quality: 153413 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 154684; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1351: contig of 1351 bp in length
            *     1352 1451: gap of      100 bp
            *     1452     4369: contig of 2918 bp in length
            *     4370 4469: gap of      100 bp
            *     4470     8873: contig of 4404 bp in length
            *     8874 8973: gap of      100 bp
            *     8974    13167: contig of 4194 bp in length
            *    13168 13267: gap of      100 bp
            *    13268    18989: contig of 5722 bp in length
            *    18990 19089: gap of      100 bp
            *    19090    24989: contig of 5900 bp in length
            *    24990 25089: gap of      100 bp
            *    25090    33460: contig of 8371 bp in length
            *    33461 33560: gap of      100 bp
            *    33561    42328: contig of 8768 bp in length
            *    42329 42428: gap of      100 bp
            *    42429    50674: contig of 8246 bp in length
            *    50675 50774: gap of      100 bp
            *    50775    59862: contig of 9088 bp in length
            *    59863 59962: gap of      100 bp
            *    59963    71699: contig of 11737 bp in length
            *    71700 71799: gap of      100 bp
            *    71800    82583: contig of 10784 bp in length
            *    82584 82683: gap of      100 bp
            *    82684    95072: contig of 12389 bp in length
            *    95073 95172: gap of      100 bp
            *    95173   108949: contig of 13777 bp in length
            *   108950 109049: gap of      100 bp
            *   109050   126668: contig of 17619 bp in length
            *   126669 126768: gap of      100 bp
            *   126769   156184: contig of 29416 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-478O20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1351
                     /note="assembly_fragment"
     misc_feature    1452. .4369
                     /note="assembly_fragment"
     misc_feature    4470. .8873
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8974. .13167
                     /note="assembly_fragment"
     misc_feature    13268. .18989
                     /note="assembly_fragment"
     misc_feature    19090. .24989
                     /note="assembly_fragment"
     misc_feature    25090. .33460
                     /note="assembly_fragment"
     misc_feature    33561. .42328
                     /note="assembly_fragment"
     misc_feature    42429. .50674
                     /note="assembly_fragment"
     misc_feature    50775. .59862
                     /note="assembly_fragment"
     misc_feature    59963. .71699
                     /note="assembly_fragment"
     misc_feature    71800. .82583
                     /note="assembly_fragment"
     misc_feature    82684. .95072
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    95173. .108949
                     /note="assembly_fragment"
     misc_feature    109050. .126668
                     /note="assembly_fragment"
     misc_feature    126769. .156184
                     /note="assembly_fragment"
BASE COUNT    51507 a  26941 c  26653 g  49578 t   1505 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 71;  Length 156184;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db 123532 GCAAGTCACAGAATTA 123517


RESULT  20
AP001390/c
LOCUS       AP001390   174815 bp    DNA             HTG       30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
            DRAFT SEQUENCE, 21 unordered pieces.
```

```
ACCESSION   AP001390
VERSION     AP001390.2  GI:8117301
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-635B11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 174,815 genomic DNA of 18q23
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7259325.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-635B11
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159194 bases at least Q40
                Consensus quality: 167087 bases at least Q30
                Consensus quality: 170693 bases at least Q20
                Insert size: 172815; sum-of-contigs
                Quality coverage: 4.94x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            21 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     58823 contig of  58823 bp in length
              58924     73243 contig of  14320 bp in length
              73344     83572 contig of  10229 bp in length
              83673     94547 contig of  10875 bp in length
              94648    104597 contig of   9950 bp in length
             104698    114347 contig of   9650 bp in length
             114448    122080 contig of   7633 bp in length
             122181    128465 contig of   6285 bp in length
             128566    135791 contig of   7226 bp in length
             135892    142934 contig of   7043 bp in length
             143035    147999 contig of   4965 bp in length
             148100    151734 contig of   3635 bp in length
             151835    155576 contig of   3742 bp in length
             155677    159193 contig of   3517 bp in length
             159294    161359 contig of   2066 bp in length
             161460    163515 contig of   2056 bp in length
             163616    166614 contig of   2999 bp in length
             166715    169056 contig of   2342 bp in length
             169157    171527 contig of   2371 bp in length
             171628    173248 contig of   1621 bp in length
             173349    174815 contig of   1467 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    58823: contig of 58823 bp in length
            *   58824 58923: gap of      100 bp
            *   58924    73243: contig of 14320 bp in length
            *   73244 73343: gap of      100 bp
            *   73344    83572: contig of 10229 bp in length
            *   83573 83672: gap of      100 bp
            *   83673    94547: contig of 10875 bp in length
            *   94548 94647: gap of      100 bp
            *   94648   104597: contig of 9950 bp in length
            *  104598 104697: gap of      100 bp
            *  104698   114347: contig of 9650 bp in length
            *  114348 114447: gap of      100 bp
            *  114448   122080: contig of 7633 bp in length
            *  122081 122180: gap of      100 bp
            *  122181   128465: contig of 6285 bp in length
            *  128466 128565: gap of      100 bp
            *  128566   135791: contig of 7226 bp in length
            *  135792 135891: gap of      100 bp
            *  135892   142934: contig of 7043 bp in length
            *  142935 143034: gap of      100 bp
            *  143035   147999: contig of 4965 bp in length
            *  148000 148099: gap of      100 bp
            *  148100   151734: contig of 3635 bp in length
            *  151735 151834: gap of      100 bp
            *  151835   155576: contig of 3742 bp in length
            *  155577 155676: gap of      100 bp
            *  155677   159193: contig of 3517 bp in length
            *  159194 159293: gap of      100 bp
            *  159294   161359: contig of 2066 bp in length
            *  161360 161459: gap of      100 bp
            *  161460   163515: contig of 2056 bp in length
            *  163516 163615: gap of      100 bp
            *  163616   166614: contig of 2999 bp in length
            *  166615 166714: gap of      100 bp
            *  166715   169056: contig of 2342 bp in length
            *  169057 169156: gap of      100 bp
            *  169157   171527: contig of 2371 bp in length
            *  171528 171627: gap of      100 bp
            *  171628   173248: contig of 1621 bp in length
            *  173249 173348: gap of      100 bp
            *  173349   174815: contig of 1467 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q23"
                     /clone="RP11-635B11"
     misc_feature    1. .58823
                     /note="assembly_fragment"
     misc_feature    58924. .73243
                     /note="assembly_fragment"
     misc_feature    73344. .83572
                     /note="assembly_fragment"
     misc_feature    83673. .94547
                     /note="assembly_fragment"
     misc_feature    94648. .104597
                     /note="assembly_fragment"
     misc_feature    104698. .114347
                     /note="assembly_fragment"
     misc_feature    114448. .122080
                     /note="assembly_fragment"
     misc_feature    122181. .128465
                     /note="assembly_fragment"
     misc_feature    128566. .135791
                     /note="assembly_fragment"
     misc_feature    135892. .142934
                     /note="assembly_fragment"
```

```
     misc_feature    143035. .147999
                     /note="assembly_fragment"
     misc_feature    148100. .151734
                     /note="assembly_fragment"
     misc_feature    151835. .155576
                     /note="assembly_fragment"
     misc_feature    155677. .159193
                     /note="assembly_fragment"
     misc_feature    159294. .161359
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    161460. .163515
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    163616. .166614
                     /note="assembly_fragment"
     misc_feature    166715. .169056
                     /note="assembly_fragment"
     misc_feature    169157. .171527
                     /note="assembly_fragment"
     misc_feature    171628. .173248
                     /note="assembly_fragment"
     misc_feature    173349. .174815
                     /note="assembly_fragment"
BASE COUNT    52310 a  30923 c  32183 g  57399 t   2000 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 82;  Length 174815;
  Best Local Similarity   100.0%;  Pred. No. 4.8;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   44384 GCAAGTCACAGAATTA 44369


RESULT  21
AC020734/c
LOCUS       AC020734   196172 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-665I14, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC020734
VERSION     AC020734.3  GI:8570370
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196172)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196172)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:7024167.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0665I14
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 85%
            Sequencing vector: plasmid; 15%
            Chemistry: Dye-primer ET; 85% of reads
            Chemistry: Dye-terminator Big Dye; 15% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 184493 bases at least Q40
            Consensus quality: 187694 bases at least Q30
            Consensus quality: 189523 bases at least Q20
            Insert size: 164000; agarose-fp
            Insert size: 193672; sum-of-contigs
            Quality coverage: 4.86 in Q20 bases; agarose-fp
            Quality coverage: 4.16 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1451: contig of 1451 bp in length
            *     1452     1551: gap of unknown length
            *     1552     3593: contig of 2042 bp in length
            *     3594     3693: gap of unknown length
            *     3694     7120: contig of 3427 bp in length
            *     7121     7220: gap of unknown length
            *     7221     9653: contig of 2433 bp in length
            *     9654     9753: gap of unknown length
            *     9754    12670: contig of 2917 bp in length
            *    12671    12770: gap of unknown length
            *    12771    15795: contig of 3025 bp in length
            *    15796    15895: gap of unknown length
            *    15896    17819: contig of 1924 bp in length
            *    17820    17919: gap of unknown length
            *    17920    20983: contig of 3064 bp in length
            *    20984    21083: gap of unknown length
            *    21084    25206: contig of 4123 bp in length
            *    25207    25306: gap of unknown length
            *    25307    28986: contig of 3680 bp in length
            *    28987    29086: gap of unknown length
            *    29087    32554: contig of 3468 bp in length
            *    32555    32654: gap of unknown length
            *    32655    37464: contig of 4810 bp in length
            *    37465    37564: gap of unknown length
            *    37565    42541: contig of 4977 bp in length
            *    42542    42641: gap of unknown length
            *    42642    47417: contig of 4776 bp in length
            *    47418    47517: gap of unknown length
            *    47518    52947: contig of 5430 bp in length
            *    52948    53047: gap of unknown length
            *    53048    58091: contig of 5044 bp in length
            *    58092    58191: gap of unknown length
            *    58192    65147: contig of 6956 bp in length
            *    65148    65247: gap of unknown length
            *    65248    71454: contig of 6207 bp in length
            *    71455    71554: gap of unknown length
            *    71555    79605: contig of 8051 bp in length
            *    79606    79705: gap of unknown length
            *    79706    88670: contig of 8965 bp in length
            *    88671    88770: gap of unknown length
            *    88771    97995: contig of 9225 bp in length
            *    97996    98095: gap of unknown length
            *    98096   109354: contig of 11259 bp in length
            *   109355   109454: gap of unknown length
            *   109455   122284: contig of 12830 bp in length
            *   122285   122384: gap of unknown length
            *   122385   139948: contig of 17564 bp in length
            *   139949   140048: gap of unknown length
            *   140049   165725: contig of 25677 bp in length
            *   165726   165825: gap of unknown length
            *   165826   196172: contig of 30347 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-665I14"
     misc_feature    1. .1451
                     /note="assembly_name:Contig4"
```

```
     misc_feature    1552. .3593
                     /note="assembly_name:Contig5"
     misc_feature    3694. .7120
                     /note="assembly_name:Contig6"
     misc_feature    7221. .9653
                     /note="assembly_name:Contig7"
     misc_feature    9754. .12670
                     /note="assembly_name:Contig8"
     misc_feature    12771. .15795
                     /note="assembly_name:Contig9"
     misc_feature    15896. .17819
                     /note="assembly_name:Contig10"
     misc_feature    17920. .20983
                     /note="assembly_name:Contig11"
     misc_feature    21084. .25206
                     /note="assembly_name:Contig12"
     misc_feature    25307. .28986
                     /note="assembly_name:Contig13"
     misc_feature    29087. .32554
                     /note="assembly_name:Contig14"
     misc_feature    32655. .37464
                     /note="assembly_name:Contig15"
     misc_feature    37565. .42541
                     /note="assembly_name:Contig16"
     misc_feature    42642. .47417
                     /note="assembly_name:Contig17"
     misc_feature    47518. .52947
                     /note="assembly_name:Contig18"
     misc_feature    53048. .58091
                     /note="assembly_name:Contig19"
     misc_feature    58192. .65147
                     /note="assembly_name:Contig20"
     misc_feature    65248. .71454
                     /note="assembly_name:Contig21"
     misc_feature    71555. .79605
                     /note="assembly_name:Contig22"
     misc_feature    79706. .88670
                     /note="assembly_name:Contig23"
     misc_feature    88771. .97995
                     /note="assembly_name:Contig24"
     misc_feature    98096. .109354
                     /note="assembly_name:Contig25
                     clone_end:T7
                     vector_side:right"
     misc_feature    109455. .122284
                     /note="assembly_name:Contig26"
     misc_feature    122385. .139948
                     /note="assembly_name:Contig27"
     misc_feature    140049. .165725
                     /note="assembly_name:Contig28"
     misc_feature    165826. .196172
                     /note="assembly_name:Contig29"
BASE COUNT    64027 a  34406 c  33812 g  61398 t   2529 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 66;  Length 196172;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   94789 GCAAGTCACAGAATTA 94774


RESULT  22
AR080743
LOCUS       AR080743     1771 bp    DNA             PAT       31-AUG-2000
DEFINITION  Sequence 48 from patent US 5968826.
ACCESSION   AR080743
VERSION     AR080743.1  GI:10007473
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1771)
  AUTHORS   Bennett,C.Frank, Condon,T.P. and Cowsert,L.M.
  TITLE     Antisense inhibition of integrin .alpha.4 expression
  JOURNAL   Patent: US 5968826-A 48 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1771
                     /organism="unknown"
BASE COUNT      371 a    471 c    519 g    410 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db    312 GCAAGTCACAGAATT 326


RESULT  23
MUSINTESUB
LOCUS       MUSINTESUB   1771 bp    DNA             ROD       18-APR-1996
DEFINITION  Mus musculus (clone A4G16) integrin alpha-4 subunit gene, 5' end
            cds.
ACCESSION   L20788
VERSION     L20788.1  GI:309416
KEYWORDS    integrin alpha-4 subunit.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1771)
  AUTHORS   De Meirsman,C., Schollen,E., Jaspers,M., Ongena,K., Matthijs,G.,
            Marynen,P. and Cassiman,J.J.
  TITLE     Cloning and characterization of the promoter region of the murine
            alpha-4 integrin subunit
  JOURNAL   DNA Cell Biol. 13 (7), 743-754 (1994)
  MEDLINE   95290094
FEATURES             Location/Qualifiers
     source          1. .1771
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="3T3"
                     /cell_type="fibroblasts"
                     /tissue_lib="NIH 3T3 in lambda-FIXII"
     GC_signal       503. .508
                     /note="putative"
     GC_signal       845. .850
     exon            952. .1389
     GC_signal       complement(1055. .1060)
     gene            join(1193. .1389,1711. .1771)
                     /gene="VLA-4"
     CDS             join(1193. .1389,1711. .>1771)
                     /gene="VLA-4"
                     /note="putative"
                     /codon_start=1
                     /product="integrin alpha-4 subunit"
                     /protein_id="AAA97501.1"
                     /db_xref="GI:309417"
                     /translation="MAAEARCRPRSRGIALREAVMLLLYFGVPTGHSYNLDPENALLY
                     QGPSGTLFGYSVVLHSHGSKRWLIVGAPTASWLSNASVVNPG"
     intron          1390. .1710
     exon            1711. .1771
                     /gene="VLA-4"
BASE COUNT      371 a    471 c    519 g    410 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 94;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 37;
```

```
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 GCAAGTCACAGAATT 326


RESULT  24
G07181/c
LOCUS       G07181       1829 bp    DNA             STS       19-OCT-1995
DEFINITION  human STS WI-9211.
ACCESSION   G07181
VERSION     G07181.1  GI:860426
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1829)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: CAGCCACACTAACTGCATTAGC
            Primer B: GACAGAAATGGATTGTATAAATGGC
            STS size: 111
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Prepared with primer pairs derived from D13644 -- Unigene.
FEATURES             Location/Qualifiers
     source          1. .1829
                     /organism="Homo sapiens"
     STS             891. .1001
     primer_bind     891. .912
     primer_bind     complement(977. .1001)
BASE COUNT      539 a    321 c    320 g    590 t     59 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 54;  Length 1829;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    1689 CAAGTCACAGAATTA 1675


RESULT  25
MMINTEGR01
LOCUS       MMINTEGR01   1949 bp    DNA             ROD       02-OCT-1996
DEFINITION  Mus musculus alpha-4 integrin gene, exons 1 and 2.
ACCESSION   U34627
VERSION     U34627.1  GI:1173582
KEYWORDS    .
SEGMENT     1 of 22
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   De Meirsman,C., Schollen,E., Jaspers,M., Ongena,K., Matthijs,G.,
            Marynen,P. and Cassiman,J.J.
  TITLE     Cloning and characterization of the promoter region of the murine
            alpha-4 integrin subunit
  JOURNAL   DNA Cell Biol. 13 (7), 743-754 (1994)
  MEDLINE   95290094
REFERENCE   2  (bases 1 to 1949)
  AUTHORS   De Meirsman,C., Jaspers,M., Schollen,E. and Cassiman,J.J.
  TITLE     The genomic structure of the murine alpha 4 integrin gene
  JOURNAL   DNA Cell Biol. 15 (7), 595-603 (1996)
  MEDLINE   96326295
REFERENCE   3  (bases 1 to 1949)
  AUTHORS   De Meirsman,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-1995) Catherine De Meirsman, Center for Human
            Genetics, University of Leuven, Herestraat 49, Leuven 3000, Belgium
FEATURES             Location/Qualifiers
     source          1. .1949
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="spleen"
                     /note="NIH-3T3 B6/CBA spleen genomic library in lambda Fix
                     II; NIH-3T3 nd B6/CBA spleen genomic library in lambda
                     FixII"
     promoter        1. .952
     exon            953. .1389
                     /number=1
     5'UTR           953. .1192
     exon            1711. .1832
                     /number=2
BASE COUNT      420 a    514 c    563 g    452 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 94;  Length 1949;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 GCAAGTCACAGAATT 326


RESULT  26
ECHK31IM
LOCUS       ECHK31IM     2340 bp    DNA             BCT       25-SEP-1998
DEFINITION  E.coli HK31IM gene.
ACCESSION   X82231
VERSION     X82231.1  GI:3660492
KEYWORDS    alpha-polypeptide; beta-polypeptide; R.EcoHK31I protein.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2340)
```

```
  AUTHORS   Lee,K.F., Kam,K.M. and Shaw,P.C.
  TITLE     A bacterial methyltransferase M.EcoHK311 requires two proteins for
            in vitro methylation
  JOURNAL   Nucleic Acids Res. 23 (1), 103-108 (1995)
  MEDLINE   95175351
REFERENCE   2  (bases 1 to 2340)
  AUTHORS   Shaw,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1994) P. Shaw, Dept. of Biochemistry, The Chinese
            University of Hong Kong, Shatin NT, HONG KONG
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 2340)
  AUTHORS   Shaw,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1997) P. Shaw, Dept. of Biochemistry, The Chinese
            University of Hong Kong, Shatin NT, HONG KONG
REFERENCE   4  (bases 1 to 2340)
  AUTHORS   Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D.
  TITLE     Sequence comparison of the EcoHK31I and EaeI
            restriction-modification systems suggests an intergenic transfer of
            genetic material
  JOURNAL   Biol. Chem. 379 (4-5), 437-441 (1998)
  MEDLINE   98290299
COMMENT     On Sep 28, 1998 this sequence version replaced gi:639971.
FEATURES             Location/Qualifiers
     source          1. .2340
                     /organism="Escherichia coli"
                     /strain="HK31"
                     /db_xref="taxon:562"
     gene            407. .1336
                     /gene="ecoHK31IM alpha"
     CDS             407. .1336
                     /gene="ecoHK31IM alpha"
                     /function="methyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="M.EcoHK31I alpha polypeptide"
                     /protein_id="CAA57706.1"
                     /db_xref="GI:3660493"
                     /db_xref="SPTREMBL:O87740"
                     /translation="MKKKPLKQYKVSSFFAGIGGFDLGLEKAGMEVVFQCEINKFCQK
                     VLRKNWSKVPLHTDITRLNADEIPESNVWCGGFPCQDVSSANQGKRKGLEGARSGLFY
                     TYAKLIEERKPEWLIIENVPGLLNSHNGQDFKVVIDTLVEFGYGVSWRVLDAKYFGTP
                     QRRRRVYIVASLGDMRSARVLFEPGATRIVDKQGQGERANASGASATGLSKADIYSIQ
                     HASIGRKATAGPQAKGYRNDGETYTLDSRGSSDAVCAPDAAFRVRASTGISRELDSNR
                     FRAVGNAVAVPIVEWLGKRIVLVDQQVLNEALS"
     gene            717. .1247
                     /gene="ecoHK31IM beta"
     CDS             717. .1247
                     /gene="ecoHK31IM beta"
                     /function="methyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="M.EcoHK31I polypeptide beta"
                     /protein_id="CAA57705.1"
                     /db_xref="GI:3660494"
                     /db_xref="SPTREMBL:Q47257"
                     /translation="MQNSSKKESLNGLLLKMFPDCSTATMDKTSKLSSIRWSNSGMAF
                     RGEYWMQNTLEHPSVEEECTLSQVLETCAPLESFLNPEQLESLINRAKERGQMLPEPL
                     LQAYQKQISILSSMQVLDEKQPQDLKQKDTGTMEKPTLLTQEEVQMLYVRRMLPSEYE
                     RLQGFPENWTLIDSEQ"
     gene            complement(1317. .2273)
                     /gene="ecoHK31IR"
     CDS             complement(1317. .2273)
                     /gene="ecoHK31IR"
                     /function="restriction endonuclease"
                     /codon_start=1
                     /transl_table=11
                     /product="R.EcoHK31I protein"
                     /protein_id="CAA57707.1"
                     /db_xref="GI:3660495"
                     /db_xref="SPTREMBL:O87741"
                     /translation="MAVSIAEYLGQRTDVDQQIVPVAKGTQIQCPFMDRTCDKASKVK
                     NPTPPVCSVRKPDGTVWIVCEHRLCSTRQKKTVIVNGRKKQIENILVEHQRDILRKVA
                     KLIYQDPELQDSEIGVRREVNIPLPDSDNSYHADYVMRNFSGRGRVDEVLLEMQGGGE
                     TSSTGEITRHIAAWADLEFPTNEILRQPVAANTIETNAWRRQQEQFLVKGNVVDQTGG
                     KIVFAVGSLLYDYLHRRFRNANLRDLKAHNWTLCILAFKEDTSDEPRPGPIPLIIDDS
                     KTLFTNYSTFVRFLTDQGAPRPELFEGSFLRLDNSSVTIMPR"
BASE COUNT      679 a    458 c    527 g    676 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 2;  Length 2340;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 acattgtcgcaagtc 16
           ||||||||||||||||
Db     906 ACATTGTCGCAAGTC 920


RESULT  27
AX098224/c
LOCUS       AX098224      3748 bp    DNA             PAT       30-MAR-2001
DEFINITION  Sequence 136 from Patent WO0118542.
ACCESSION   AX098224
VERSION     AX098224.1  GI:13515348
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3748)
  AUTHORS   Lee,J., Thompsho,P. and Lillie,J.
  TITLE     Identification, assessment, prevention, and therapy of ovarian
            cancer
  JOURNAL   Patent: WO 0118542-A 136 15-MAR-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3748
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1110 a    825 c    761 g   1052 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 3748;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    2627 CAAGTCACAGAATTA 2613


RESULT  28
D13644/c
LOCUS       D13644        4602 bp    mRNA            PRI       10-JUL-1997
DEFINITION  Human mRNA for KIAA0019 gene, complete cds.
ACCESSION   D13644
VERSION     D13644.1  GI:1531551
KEYWORDS    KIAA0019.
SOURCE      Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4602)
  AUTHORS   Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1992) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:0438-52-3930, Fax:0438-52-3931)
REFERENCE   2  (bases 1 to 4602)
  AUTHORS   Nomura,N., Miyajima,N., Kawarabayasi,Y. and Tabata,S.
  TITLE     Prediction of new human genes by entire sequencing of randomly
```

```
            sampled cDNA clones
  JOURNAL   Unpublished (1994)
REFERENCE   3  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1
  JOURNAL   DNA Res. 1 (1), 27-35 (1994)
  MEDLINE   96051387
  REMARK    Erratum:[[published erratum appears in DNA Res 1995 Aug
            31;2(4):211]]
REFERENCE   4  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1 (supplement)
  JOURNAL   DNA Res. 1 (1), 47-56 (1994)
  MEDLINE   96051389
REFERENCE   5  (bases 1 to 4602)
  AUTHORS   Matoskova,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N.,
            Robbins,K.C. and Di Fiore,P.P.
  TITLE     RN-tre identifies a family of tre-related proteins displaying a
            novel potential protein binding domain
  JOURNAL   Oncogene 12 (12), 2563-2571 (1996)
  MEDLINE   96293402
COMMENT     On Sep 10, 1996 this sequence version replaced gi:286016.
FEATURES             Location/Qualifiers
     source          1. .4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="KG-1"
                     /cell_type="myoblast"
                     /chromosome="10"
                     /map="10p13"
                     /sex="male"
     5'UTR           <1. .279
     gene            280. .2766
                     /gene="KIAA0019"
     CDS             280. .2766
                     /gene="KIAA0019"
                     /codon_start=1
                     /product="protein related N-ternimus of tre oncogene"
                     /protein_id="BAA02807.1"
                     /db_xref="GI:2104571"
                     /translation="MNSDQDVALKLAQERAEIVAKYDRGREGAEIEPWEDADYLVYKV
                     TDRFGFLHEEELPDHNVAVERQKHLEIERTTKWLKMLKGWEKYKNTEKFHRRIYKGIP
                     LQLRGEVWALLLEIPKMKEETRDLYSKLKHRARGCSPDIRQIDLDVNRTFRDHIMFRD
                     RYGVKQQSLFHVLAAYSIYNTEVGYCQGMSQITALLLMYMNEEDAFWALVKLFSGPKH
                     AMHGFFVQGFPKLLRFQEHHEKILNKFLSKLKQHLDSQEIYTSFYTMKWFFQCFLDRT
                     PFTLNLRIWDIYIFEGERVLTAMSYTILKLHKKHLMKLSMEELVEFFQETLAKDFFFE
                     DDFVIEQLQISMTELKRAKLDLPEPGKEDEYPKKPLGQLPPELQSWGVHHLSNGQRSV
                     GRPSPLASGRRESGAPHRRHEHSPHPQSRTGTPERAQPPRRKSVEEESKKLKDEADFQ
                     RKLPSGPQDSSRQYNHAAANQNSNATSNIRKEFVPKWNKPSDVSATERTAKYTMEGKG
                     RAAHPALAVTVPGPAEVRVSNVRPKMKALDAEDGKRGSTASQYDNVPGPELDSGASVE
                     EALERAYSQSPRHALYPPSPRKHAEPSSSPSKVSNKFTFKVQPPSHARYPSQLDGEAR
                     GLAHPPSYSNPPVYHGNSPKHFPTANSSFASPQFSPGTQLNPSRRPHGSTLSVSASPE
                     KSYSRPSPLVLPSSRIEVLPVDTGAGGYSGNSGSPKNGKLIIPPVDYLPDNRTWSEVS
                     YTYRPETQGQSWTRDASRGNLPKYSAFQLAPFQDHGLPAVSVDSPVRYKASPAAEDAS
                     PSGYPYSGPPPPAYHYRNRDGLSIQESVLL"
     3'UTR           2767. .4602
BASE COUNT     1340 a    1041 c    999 g    1222 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 91;  Length 4602;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4462 CAAGTCACAGAATTA 4448


RESULT  29
AF039582
LOCUS       AF039582     4636 bp    DNA             BCT       05-DEC-2000
DEFINITION  Enterobacter aerogenes putative integrase (int(p)) gene, partial
            cds; and EaeI methyltransferase alpha subunit (eaeIM-a), EaeI
            methyltransferase beta subunit (eaeIM-b), and EaeI restriction
            endonuclease (eaeIR) genes, complete cds.
ACCESSION   AF039582
VERSION     AF039582.1  GI:2760953
KEYWORDS    .
SOURCE      Enterobacter aerogenes.
  ORGANISM  Enterobacter aerogenes
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Enterobacter.
REFERENCE   1  (bases 1 to 4636)
  AUTHORS   Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D.
  TITLE     Sequence comparison of the EcoHK31I and EaeI
            restriction-modification systems suggests an intergenic transfer of
            genetic material
  JOURNAL   Biol. Chem. 379 (4-5), 437-441 (1998)
  MEDLINE   98290299
   PUBMED   9628335
REFERENCE   2  (bases 1 to 4636)
  AUTHORS   Lunnen,K.D., Picone,S.J. and Wilson,G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1997) Research, New England Biolabs, 32 Tozer
            Road, Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1. .4636
                     /organism="Enterobacter aerogenes"
                     /db_xref="taxon:548"
     gene            <1. .1035
                     /gene="int(p)"
     CDS             <1. .1035
                     /gene="int(p)"
                     /note="similar to phage P4 integrase SP:P39347"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integrase"
                     /protein_id="AAB95339.1"
                     /db_xref="GI:2760957"
                     /translation="KLVAAGIDPREHKRAVKEEQTKGAITFESVAREWHAANKKWTEE
                     HSRRVLKSLEDNLFPAIGKRSIDSFSTRDLLVPIKVVEATGRLEVASRLQQRTNAIMR
                     YAVQSGLIDYNPAQEMAGAVASSNRVHRPALELKRLPELLQRIDSYSGRPLTRLAVEL
                     TLLIFIRSSELRFARWPEIDFERAMWTIPAEREAIEGVKHSQRGSKMRTPHLVPLSRQ
                     ALEILKQVHKFSGERDFVFVGDHNPRKPMSENTVNKALRVMGYDTKVEVCGHGFRTMA
                     CSSLIESGLWSRDAVERQMSHMERNSVRAAYIHKAEHLDERKLMLQWWADFLDANRQK
                     GISPFDYAKA"
     gene            1228. .2157
                     /gene="eaeIM-a"
     CDS             1228. .2157
                     /gene="eaeIM-a"
                     /function="converts YGGCCR to YGGmCCR"
                     /note="M.EaeIa"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI methyltransferase alpha subunit"
                     /protein_id="AAB95336.1"
                     /db_xref="GI:2760954"
                     /translation="MKKKPLKQYKVSSFFAGIGGFDLGLEKAGMEVVFQCEINKFCQK
                     VLRKNWPNVPLHTDITQLNADEIPESSVWCGGFPCQDVSSANQGKRKGLEGARSGLFY
                     TYAKLIEERKPEWLVIENVPGLLNSHNGQDFKVVIDTLVELGYGVSWRVLDAKYFGTP
                     QRRRRVYIVASLGDMRSARVLFEPGATRIVDKQGQGERTNASGASDTGLSKADIYSIQ
                     HASIGRKATAGPQAKGYRNDGETYTLDSRGSSDAVCAPDAAFRVRASAGISRELDSNR
                     FRAVGNAVAVPIVEWIGKRIVLVDQQMLNEALS"
     gene            1538. .2068
                     /gene="eaeIM-b"
     CDS             1538. .2068
                     /gene="eaeIM-b"
```

```
                     /note="M.EaeIb; translated from 3'-half of the eaeIM-a
                     gene in an alternative reading frame"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI methyltransferase beta subunit"
                     /protein_id="AAB95337.1"
                     /db_xref="GI:2760955"
                     /translation="MQNSSKKENLNGLLLKMFPDCSTATMDKTSKLSSIRWSNSGMAF
                     RGEYWMQNTLEHPSVEEECTLSQVLETCAPLESFLNPEQLESLINRAKERGQMLPEPL
                     IQAYQKQISILSSMQVLDEKQPQDLKQKDTGTTEKPTHLTQEEVQMLYVRRMLPSEYE
                     RLQGFPENWTLIDSEQ"
     gene            complement(2138. .3085)
                     /gene="eaeIR"
     CDS             complement(2138. .3085)
                     /gene="eaeIR"
                     /function="cleaves between Y and G in YGGCCR"
                     /note="R.EaeI; unable to cleave methylated ds sequence
                     YGGmCCR"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI restriction endonuclease"
                     /protein_id="AAB95338.1"
                     /db_xref="GI:2760956"
                     /translation="MAVSIAEYLGQRTDVDQQIVPVSKGTLAQCPFMDRTCDKVSKNN
                     HPVCSVRKPDSTIWIVCEHRLCSTRQKKTVISGERTQQVDNVLVEHQRDILRKVAKLI
                     YQDPELQDSEIGVKREVNIPLPDSDNSYHADYVMRNFSGRGRVDEVLLEMQGGGETSS
                     TGAITRHIAAWADLEFPTNEILRRPVAANTIETNAWRRQQEQFLVKGNVVDQTGGKIV
                     FAVGSLLYDYLHRRFRNANLRDLKAHNWTLCILAFKEDTSDEPQPGPIPLIIDDSKTL
                     FTNYSTFVRFLTDQGAPRPELFEGSFLRLDNSTVTIMPR"
BASE COUNT     1362 a    893 c   1074 g   1307 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 1;  Length 4636;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 acattgtcgcaagtc 16
          |||||||||||||||
Db   1727 ACATTGTCGCAAGTC 1741
```

```
RESULT  30
AF152313/c
LOCUS       AF152313     5218 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 5 (PCDH-alpha5) mRNA, complete
            cds.
ACCESSION   AF152313
VERSION     AF152313.1  GI:5456913
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5218
                     /gene="PCDH-alpha5"
     CDS             1. .2811
                     /gene="PCDH-alpha5"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 5"
                     /protein_id="AAD43707.1"
                     /db_xref="GI:5456914"
                     /translation="MVYSRRGSLGSRLLLLWLLLAYWKAGSGQLHYSIPEEAKHGTFV
                     GRIAQDLGLELAELVPRLFRVASKGRGDLLEVNLQNGILFVNSRIDREELCRRRAECS
                     IHLEVIVDRPLQVFHVEVAVKDINDNPPRFSRQEQRLFILESRMPDSRFPLEGASDLD
                     IGANAQLRYRLNPNEYFDLDVKTNEEETNFLELVLRKSLDREETQEHRLLVIATDGGK
                     PELTGTVQLLINVLDANDNAPEFDKSIYNVRLLENAPSGTLVIKLNASDADEGINKEI
                     VYFFSNLVLDDVKSKFIINSNTGEIKVNGELDYEDYNSYEINIDAMDKSTFPLSGHCK
                     VVVKLLDVNDNTPEMAITTLFLPVKEDAPLSTVIALISVSDRDSGANGQVTCSLMPHV
                     PFKLVSTFKNYYSLVLDSALDRESVSVYELVVTARDGGSPSLWATASVSVEVADVNDN
                     APAFAQPQYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERPLSSYVS
                     VHAESGKVYALQPLDHEEVELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALLV
                     PRVGGTGGAVSELVPRSVGAGHVVAKVRAVDPDSGYNAWLSYELQPAPGSARIPFRVG
                     LYTGEISTTRSLDETEAPRHRLLVLVKDHGEPPLTATATVLVSLVESGQAPKASSRAS
                     AGAVGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAQPTEAVCTRGKPTLLC
                     SSAVGSWSYSQQRRQRVCSGEAPPKTDLMAFSPSLPQGPTSTDNPRQPNPDWRYSASL
                     RAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSNSWTFK
                     YGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETKKKKKK
                     KKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1385 a    1207 c   1281 g   1345 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 89;  Length 5218;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 caagtcacagaatta 25
          |||||||||||||||
Db   4115 CAAGTCACAGAATTA 4101
```

```
RESULT  31
AF152315/c
LOCUS       AF152315     5221 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 7 (PCDH-alpha7) mRNA, complete
            cds.
ACCESSION   AF152315
VERSION     AF152315.1  GI:5456917
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .5221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5221
                     /gene="PCDH-alpha7"
     CDS             1. .2814
                     /gene="PCDH-alpha7"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 7"
                     /protein_id="AAD43709.1"
                     /db_xref="GI:5456918"
                     /translation="MVCPNGYDPGGRHLLLFIIILAAWEAGRGQLHYSVPEEAKHGNF
                     VGRIAQDLGLELAELVPRLFRAVCKFRGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVERPLQVFHVDVEVKDINDNPPVFPATQRNLFIAESRPLDSRFPLEGASDA
                     DIGENALLTYRLSPNEYFFLDVPTSNQQVKPLGLVLRKLLDREETPELHLLLTATDGG
                     KPELTGTVQLLITVLDNNDNAPVFDRTLYTVKLPENVSIGTLVIHPNASDLDEGLNGD
                     IIYSFSSDVSPDIKSKFHMDPLSGAITVIGHMDFEESRAHKIPVEAVDKGFPPLAGHC
                     TVLVEVVDVNDNAPQLTLTSLSLPIPEDAQPGTVITLISVFDRDFGVNGQVTCSLTPR
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAGDADAQKNALVSYSLVELRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTGGAVRELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPVAAGASIPFRV
                     GLYTGEISTTRALDETDAPRHRLLVLVKDHGEPSLTATATVLVSLVESGQAPKASSRA
                     SLGIAGPETELVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPSSEGACSLVKPTLV
                     CSSAVGSWSFSQQRRQRVCSGEGPPKTDLMAFSPSLPQGPSSTDNPRQPNPDWRYSAS
                     LRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSNSWTF
                     KYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETKKKKK
                     KKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT      1333 a   1242 c   1295 g   1351 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 89;  Length 5221;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4118 CAAGTCACAGAATTA 4104
```

```
RESULT  32
AF152308/c
LOCUS       AF152308     5233 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 12 (PCDH-alpha12) mRNA, complete
            cds.
ACCESSION   AF152308
VERSION     AF152308.1  GI:5456901
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .5233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5233
                     /gene="PCDH-alpha12"
     CDS             1. .2826
                     /gene="PCDH-alpha12"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 12"
                     /protein_id="AAD43702.1"
                     /db_xref="GI:5456902"
                     /translation="MVIIGPRGPGSQRLLLSLLLLAAWEVGSGQLHYSVYEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREKLCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVKDINDNPPVFREREQKVPVSESAPLDSHFPLEGASDA
                     DIGVNSLLTYALSLNENFELKIKTKKDKSILPELVLRKLLDREQTPKLNLLLMVIDGG
                     KPELTGSVQIQITVLDVNDNGPAFDKPSYKVVLSENVQNDTRVIQLNASDPDEGLNGE
                     ISYGIKMILPVSEKCMFSINPDTGEIRIYGELDFEENNAYEIQVNAIDKGIPSMAGHS
                     MVLVEVLDVNDNVPEVMVTSLSLPVQEDAQVGTVIALISVSDRDSGANGQVICSLTPH
                     VPFKLVSTYKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATARVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQKNALVSYSLVERRVGEHALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     ATPAGSAGGAVSELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAAVGAHIPFHV
                     GLYTGEISTTRILDEADAPRHRLLVLVKDHGEPALTSTATVLVSLVENGQAPKTSSRA
                     SVGAVDPEAALVDINVYLIIAICAVSSLLVLTLLLYTALRCSAPPTVSRCAPGKPTLV
                     CSSAVGSWSYSQQRRQRVCSAESPPKTDLMAFSPSLQLSREDCLNPPSEPRQPNPDWR
                     YSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSN
                     SWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETK
                     KKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT      1364 a   1213 c   1305 g   1351 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 89;  Length 5233;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4130 CAAGTCACAGAATTA 4116
```

```
RESULT  33
AF152312/c
LOCUS       AF152312     5251 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 4 (PCDH-alpha4) mRNA, complete
            cds.
ACCESSION   AF152312
VERSION     AF152312.1  GI:5456911
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5251
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5251
                     /gene="PCDH-alpha4"
     CDS             1. .2844
                     /gene="PCDH-alpha4"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 4"
                     /protein_id="AAD43706.1"
                     /db_xref="GI:5456912"
                     /translation="MEFSWGSGQESRRLLLLLLLLAAWEAGNGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKGRGGLLEVNLQNGILFVNSRIDREELCRRSAEC
                     SIHLEVIVDRPLQVFHVDVEVRDINDNPPVFPATQKNLSIAESRPLDSRFPLEGASDA
                     DIGENALLTYRLSPNEYFSLEKPPDDELVKGLGLILRKSLDREEAPEIFLVLTATDGG
                     KPELTGTVQLLITVLDANDNAPAFDRTIYKVRLLENVPNGTLVIKLNASDLDEGLNGD
                     IVYSFSNDISPNVKSKFHIDPITGQIIVKGYIDFEESKSYEIIVEGIDKGQLPLSGHC
                     RVIVEVEDNNDNVPDLEFKSLSLPIREDAPLGTVIALISVSDKDMGVNGLVTCSLTSH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVTARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRAGGTGGAVSELVPWSVGVGHVVAKVRAVDADSGYNAWLSYELQPGTGGARIPFRV
                     GLYTGEISTTRALDETDAPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     LVGAVGPDAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSALPTEGACAPGKPTLV
                     CSSAVGSWSYSQQRRPRVCSGEGPPKTDLMAFSPSLPDSRDREDQLQTTEESFAKPRQ
                     PNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVG
                     AGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFG
                     KKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1360 a   1229 c   1304 g   1358 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5251;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4148 CAAGTCACAGAATTA 4134


RESULT  34
AF152306/c
LOCUS       AF152306     5254 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 10 (PCDH-alpha10) mRNA, complete
            cds.
ACCESSION   AF152306
VERSION     AF152306.1  GI:5456897
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5254
                     /gene="PCDH-alpha10"
     CDS             1. .2847
                     /gene="PCDH-alpha10"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 10"
                     /protein_id="AAD43700.1"
                     /db_xref="GI:5456898"
                     /translation="MVSRCSCLGVQCLLLSLLLLAAWEVGSGQLHYSVYEEARHGTFV
                     GRIAQDLGLELAELVQRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSVECS
                     IHLEVIVDRPLQVFHVDVEVKDINDNPPRFSVTEQKLSIPESRLLDSRFPLEGASDAD
                     VGENALLTYKLSPNEYFVLDIINKKDKDKFPVLVLRKLLDREENPQLKLLLTATDGGK
                     PEFTGSVSLLILVLDANDNAPIFDRPVYEVKMYENQVNQTLVIRLNASDSDEGINKEM
                     MYSFSSLVPPTIRRKFWINERTGEIKVNDAIDFEDSNTYEIHVDVTDKGNPPMVGHCT
                     VLVELLDENDNSPEVIVTSLSLPVKEDAQVGTVIALISVSDHDSGANGQVTCSLTPHV
                     PFKLVSTYKNYYSLVLDSALDRERVSAYELVVTARDGGSPPLWATASVSVEVADVNDN
                     APAFAQSEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRLGERSLSSYVS
                     VHAESGKVYALQPLDHEELELLQFQVSARDGGVPPLGSNLTLQVFVLDENDNAPALLA
                     SPAGSAGGAVSELVLRSVVAGHVVAKVRAVDADSGYNAWLSYELQSAAVGARIPFRVG
                     LYTGEISTTRALDETDSPRQRLLVLVKDHGEPSLTATATVLVSLVEGSQAPKASSRAS
                     VGVAPEVALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAAPTEGACGPVKPTLVCS
                     SAVGSWSYSQQRRQRVCSGEGLPKADLMAFSPSLPPCPMVDVDGEDQSIGGDHSRKPR
                     QPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPV
                     GAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITF
                     GKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1358 a   1210 c   1317 g   1369 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5254;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4151 CAAGTCACAGAATTA 4137


RESULT  35
AF152310/c
LOCUS       AF152310     5254 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 2 (PCDH-alpha2) mRNA, complete
            cds.
ACCESSION   AF152310
VERSION     AF152310.1  GI:5456905
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
```

```
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5254
                     /gene="PCDH-alpha2"
     CDS             1. .2847
                     /gene="PCDH-alpha2"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 2"
                     /protein_id="AAD43704.1"
                     /db_xref="GI:5456906"
                     /translation="MASSIRRGRGAWTRLLSLLLLAAWEVGSGQLRYSVPEEAKHGTF
                     VGRIAQDLGLELEELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHVEVIVDRPLQVFHVEVEVKDINDNPPIFPMTVKTIRFPESRLLDSRFPLEGASDA
                     DIGVNALLSYKLSSSEFFFLDIQANDELSESLSLVLGKSLDREETAEVNLLLVATDGG
                     KPELTGTVQILIKVLDVNDNEPTFAQSVYKVKLLENTANGTLVVKLNASDADEGPNSE
                     IVYSLGSDVSSTIQTKFTIDPISGEIRTKGKLDYEEAKSYEIQVTATDKGTPSMSGHC
                     KISLKLVDINDNTPEVSITSLSLPISENASLGTVIALITVSDRDSGTNGHVTCSLTPH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATTSVSIEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEEVELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRAGTAAGAVSELVPWSVGAGHVVAKVRAVDADSGYNAWLSYELQLGTGSARIPFRV
                     GLYTGEISTTRALDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     WVGAAGSEATLVDVNVYLIIAICAVSSLLVLTVLLYTALRCSVPPTEGARAPGKPTLV
                     CSSAVGSWSYSQQRRQRVCSGEDPPKTDLMAFSPSLSQGPDSAEEKQLSESEYVGKPR
                     QPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPV
                     GAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITF
                     GKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1375 a   1234 c   1300 g   1345 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 89;  Length 5254;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4151 CAAGTCACAGAATTA 4137
```

```
RESULT  36
AF152307/c
LOCUS       AF152307     5257 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 11 (PCDH-alpha11) mRNA, complete
            cds.
ACCESSION   AF152307
VERSION     AF152307.1  GI:5456899
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
```

```
REFERENCE   2  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5257
                     /gene="PCDH-alpha11"
     CDS             1. .2850
                     /gene="PCDH-alpha11"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 11"
                     /protein_id="AAD43701.1"
                     /db_xref="GI:5456900"
                     /translation="MFGFQRRGLGTPRLQLWLLLLEFWEVGSGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVQRLFRVASKTHGDLLEVNLQNGILFVNSRIDREELCGQSAEC
                     SIHLEVIVDRPLQVFHVNVEVKDINDNPPVFSLREQKLLIAESKQSDSRFPLEGASDA
                     DIEENALLTYRLSKNEYFSLDSPTNGKQIKRLSLILKKSLDREKTPELNLLLTATDGG
                     KPELTGTVRLLVQVLDVNDNDPEFDKSEYKVSLMENAAKETLVLKLNATDRDEGVNGE
                     VTYSLMSIKPNGRHLFTLDQNNGEVRVNGTLDYEENKFYKIEVQATDKGTPPMAGHCT
                     VWVEILDTNDNSPEVAVTSLSLPVREDAQPSTVIALISVSDRDSGVNGQVTCSLTPHV
                     PFKLVSTFKNYYSLVLDSALDRENVWAYELVVTARDGGSPSLWATARVSVEVADVNDN
                     APAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRLGDRALSSYVS
                     VHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLSSNVTLQVFVLDENDNAPALLA
                     TQAGSAGGAVNKLVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAAGGSRIPFRVG
                     LYTGEISTTRALDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRTL
                     AGAASPEAALVDVNVYLIIAICVVSSLLVLTLLLYTALWWSATPTEGACAPGKPTLVC
                     SRAVGSWSYSQQRRQRVCSEEGPPKTDLMAFSPSLPLGLNKEEEGERQEPGSNHPGQP
                     RQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPP
                     VGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFIT
                     FGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1393 a   1224 c   1317 g   1323 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 89;  Length 5257;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4154 CAAGTCACAGAATTA 4140
```

```
RESULT  37
AF152305/c
LOCUS       AF152305     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 1 (PCDH-alpha1) mRNA, complete
            cds.
ACCESSION   AF152305
VERSION     AF152305.1  GI:5456895
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
```

```
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha1"
     CDS             1. .2853
                     /gene="PCDH-alpha1"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 1"
                     /protein_id="AAD43699.1"
                     /db_xref="GI:5456896"
                     /translation="MVFSRRGGLGARDLLLWLLLLAAWEVGSGQLHYSIPEEAKHGTF
                     VGRVAQDLGLELAELVPRLFRVASKTHRDLLEVNLQNGILFVNSRIDREELCQWSAEC
                     SIHLELIADRPLQVFHVEVKVKDINDNPPVFRGREQIIFIPESRLLNSRFPIEGAADA
                     DIGANALLTYTLSPSDYFSLDVEASDELSKSLWLELRKYLDREETPELHLLLTATDGG
                     KPELQGTVELLITVLDVNDNAPLFDQAVYRVHLLETTANGTLVTTLNASDADEGVNGE
                     VVFSFDSGISRDIQEKFKVDSSSGEIRLIDKLDYEETKSYEIQVKAVDKGSPPMSNHC
                     KVLVKVLDVNDNAPELAVTSLYLPIREDAPLSTVIALITVSDRDSGANGQVTCSLMPH
                     VPFKLVSTFKNYYSLVLDSALDRESLSVYELVVTARDGGSPSLWATARVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSNYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTIGAVSELVPRLVGAGHVVAKVRAVDADSGYNAWLSYELQPAAGGARIPFRV
                     GLYTGEISTTRVLDEADLSRYRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     SVGVAGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSVPPTEGAYVPGKPTLV
                     CSSALGSWSNSQQRRQRVCSSEGPPKTDLMAFSPGLSPSLNTSERNEQPEANLDLSGN
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1365 a   1217 c   1306 g   1372 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  38
AF152309/c
LOCUS       AF152309     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 13 (PCDH-alpha13) mRNA, complete
            cds.
ACCESSION   AF152309
VERSION     AF152309.1  GI:5456903
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha13"
     CDS             1. .2853
                     /gene="PCDH-alpha13"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 13"
                     /protein_id="AAD43703.1"
                     /db_xref="GI:5456904"
                     /translation="MLSSWQGGPRPRQLLLWLLILAAWETGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVEVKVRDINDNPPIFPESKKRIIIAESRPPETRFPLDGASDA
                     DIGVNSALTYRLDPNDYFTLDAQNSLEQMSSLSLVLRKTLDREEIQEHSLLLTASDGG
                     KPELTGTVQLLITILDVNDNAPEFYQSVYKVTVLENAFNGTLVIKLNATDPDDGTNGD
                     IVYSFRRPVWPAVVYAFTINPNNGEIRTKGKLDFEEKKLYEISVEAVDKGNIPMAGHC
                     TLLVEVLDVNDNAPEVTITSLSLPIREDTQPSAIIALISVSDRDSGSNGQVTCTLTPH
                     VPFKLVSTYKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVGVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAQDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDSGVPPLGSNVTLQVFVLDENDNAPALL
                     TPGAGSAGGTVSELMPRSVGAGHVVAKVRAVDADSGYNAWLSYELQLAAVGARIPFRV
                     GLYTGEISTTRPLDEVDAPHHRLLVLVKDHGEPALTATATVLLSLVESGQAPQASSRA
                     SAGAVGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPPTEGACAPGKPTLV
                     CSSAAGSWSYSQQRRPRVCSGEGPHKTDLMAFSPSLPPCLGSAEGTGQREEDSECLKE
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1376 a   1234 c   1315 g   1335 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  39
AF152311/c
LOCUS       AF152311     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 3 (PCDH-alpha3) mRNA, complete
            cds.
ACCESSION   AF152311
VERSION     AF152311.1  GI:5456909
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
```

```
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha3"
     CDS             1. .2853
                     /gene="PCDH-alpha3"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 3"
                     /protein_id="AAD43705.1"
                     /db_xref="GI:5456910"
                     /translation="MLFSWREDPGAQCLLLSLLLLAASEVGSGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVEVEVKDINDNAPVFPMAVKNLFISESRQPGSRFSLEGASDA
                     DIGTNSLLTYSLDSTEYFTLDVKRNDEEIKSLGLVLKKNLNREDTPKHYLLITAIDGG
                     KPELTGTTQLKITVLDVNDNAPAFERTIYKVRLLENAPNGTLVVTVNATDLDEGVNKD
                     IAYSFNTDMSADILSKFHLDPVNGQISVKGNIDFEESKSYEIQVEATDKGNPPMSDHC
                     TVLLEIVDINDNVPELVIQSLSLPVLEDSPLSTVIALISVSDRDSGVNGQVTCSLTPH
                     VPFKLVSTFKNYYSLVLDSPLDRESVSAYELVVTARDGGSPSLWATASVSVEVADVND
                     NAPAFSQSEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     MPRVGGIGGAVSELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPGTGGARIPFRV
                     GLYTGEISTTRALDEVDAPRHRLLVLVKDHGEPSLTATATVLVSLVESGQAPKASSQA
                     SAGATGPEAALVDVNVYLIVAICAVSSLLVLTLLLYTALRCSAPPTEGDCGPGKPTLV
                     CSSAVGSWSYSQQRQQRVCSGEGLPKTDLMAFSPSLPPCPISRDREEKQDVDVDLSAK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1379 a   1225 c   1293 g   1363 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  40
AF152314/c
LOCUS       AF152314     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 6 (PCDH-alpha6) mRNA, complete
            cds.
ACCESSION   AF152314
VERSION     AF152314.1  GI:5456915
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha6"
     CDS             1. .2853
                     /gene="PCDH-alpha6"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 6"
                     /protein_id="AAD43708.1"
                     /db_xref="GI:5456916"
                     /translation="MVFTPEDRLGKQCLLLPLLLLAAWKVGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRMASKDREDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVRDINDNPPLFPVEEQRVLIYESRLPDSVFPLEGASDA
                     DVGSNSILTYKLSSSEYFGLDVKINSDDNKQIGLLLKKSLDREEAPAHNLFLTATDGG
                     KPELTGTVQLLVTVLDVNDNAPTFEQSEYEVRIFENADNGTTVIRLNASDRDEGANGA
                     ISYSFNSLVAAMVIDHFSIDRNTGEIVIRGNLDFEQENLYKILIDATDKGHPPMAGHC
                     TVLVRILDKNDNVPEIALTSLSLPVREDAQFGTVIALISVNDLDSGANGQVNCSLTPH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASLSVEVADMND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSSYI
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTGGAVSELVPRSLGAGQVVAKVRAVDADSGYNAWLSYELQPPASSARFPFRV
                     GLYTGEISTTRVLDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     SVGAAGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPPTEGACTADKPTLV
                     CSSAVGSWSYSQQRRQRVCSGEGPPKMDLMAFSPSLSPCPIMMGKAENQDLNEDHDAK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1366 a   1230 c   1315 g   1349 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  41
AF152316/c
LOCUS       AF152316     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 8 (PCDH-alpha8) mRNA, complete
            cds.
ACCESSION   AF152316
VERSION     AF152316.1  GI:5456919
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha8"
     CDS             1. .2853
                     /gene="PCDH-alpha8"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 8"
                     /protein_id="AAD43710.1"
                     /db_xref="GI:5456920"
                     /translation="MDYHWRGELGSWRLLLLLLLLAAWKVGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHRDLLEVSLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVKDVNDNPPVFRVKDQKLFVSESRMPDSRFPLEGASDA
                     DVGANSVLTYRLSSHDYFMLDVNSKNDENKLVELVLRKSLDREDAPAHHLFLTATDGG
                     KPELTGTVQLLVTVLDVNDNAPTFEQSEYEVRIFENADNGTTVIKLNASDPDEGANGA
                     ISYSFNSLVETMVIDHFSIDRNTGEIVIRGNLDFEQENLYKILIDATDKGHPPMAGHC
                     TVLVRILDKNDNVPEIALTSLSLPVREDAQFGTVIALISVNDLDSGANGQVTCSLMPH
                     VPFKLVSTFKNYYSLVLDSALDRERVSAYELVVTARDGGSPSLWATASLSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERSLSSYI
                     SVHTESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     EPRVGGTGGAASKLVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAASSPRIPFRV
                     GLYTGEISTTRVLDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRQ
                     SAGVLGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSALPTEGGCRAGKPTLV
                     CSSAVGSWSYSQQQPQRVCSGEGPPKTDLMAFSPCLPPDLGSVDVGEEQDLNVDHGLK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1357 a   1236 c   1310 g   1357 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  42
AF152317/c
LOCUS       AF152317     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 9 (PCDH-alpha9) mRNA, complete
            cds.
ACCESSION   AF152317
VERSION     AF152317.1  GI:5456921
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha9"
     CDS             1. .2853
                     /gene="PCDH-alpha9"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 9"
                     /protein_id="AAD43711.1"
                     /db_xref="GI:5456922"
                     /translation="MLYSSRGDPEGQPLLLSLLILAMWVVGSGQLHYSVPEEAEHGTF
                     VGRIAQDLGLELAELVPRLFQLDSKGRGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVKDINDNPPVFPATQKNLFIAESRPLDSRFPLEGASDA
                     DIGENALLTYRLSPNEYFFLDVPTSNQQVKPLGLVLRKLLDREETPELHLLLTATDGG
                     KPELTGTVQLLITVLDNNDNAPVFDRTLYTVKLPENVSIGTLVIHPNASDLDEGLNGD
                     IIYSFSSDVSPDIKSKFHMDPLSGAITVIGHMDFEESRAHKIPVEAVDKGFPPLAGHC
                     TLLVEVVDVNDNAPQLTIKTLSVPVKEDAQLGTVIALISVIDLDADANGQVTCSLTPH
                     VPFKLVSTYKNYYSLVLDRALDRESVSAYELVVTARDGGSPSLWATARVSVEVADVND
                     NAPAFAQSEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRLGERSLSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     TPRMRGTDGAVSEMVLRSVGAGVVVGKVRAVDADSGYNAWLSYELQPETASASIPFRV
                     GLYTGEISTTRALDETDAPRQRLLVLVKDHGEPALTATATVLVSLVESGQAPKSSSRA
                     SVGATGPEVTLVDVNVYLIIAICAVSSLLVLTLLLYTVLRCSAMPTEGECAPGKPTLV
                     CSSAVGSWSYSQQRRQRVCSGEGKQKTDLMAFSPGLSPCAGSTERTGEPSASSDSTGK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1354 a   1257 c   1320 g   1329 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  43
AF152303/c
LOCUS       AF152303     5299 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha C1 (PCDH-alpha-C1) mRNA, complete
            cds.
```

```
ACCESSION   AF152303
VERSION     AF152303.1  GI:5456891
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5299)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5299)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5299)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5299
                     /gene="PCDH-alpha-C1"
     CDS             1. .2892
                     /gene="PCDH-alpha-C1"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha C1"
                     /protein_id="AAD43697.1"
                     /db_xref="GI:5456892"
                     /translation="MVGCGVAVLCLWVSCGAAAGQLEYSVPEETERGVAVGNLSADLR
                     LPAAAMSSRNFRFLSSHRELYFGVDLPSGNLVVREPADREQLCRAKAACVLTYDLVLE
                     DPLELHKIRIHVLDTNDNSPLFPAGDVQLHIPEFLTPGARFTLPNAQDDDEGSNGILS
                     YSLSPSQHFRLDMGSRVDGSEYPELVLEKALDREQRATHLLVLTARDGGLPARSGDAQ
                     VTIIVVDTNDNAPVFERSVYRTKVPETAPNGTVLFRVQALDPDEGSNGEVQYSLSNST
                     QAELRHRFHVHPKSGEVQVAASLGPPETLLEAYIEARDEGVFGLASTAKLLVEVTDVN
                     DHAPELDFLTLSNPVPEDAAPGTVIALFSVKDEDLDSNGRVICGMSSAGPFQLTASFD
                     NYYSLLIDGPLDREQISEYQVLITASDSGSPPLSTRRTITVSVADVNDNTPNFPQPQQ
                     ELFVAENNGPGASLGRVFAQDPDLGKNGLVSYELLDVISEGPSASSLVAVESSSGAIT
                     AKTSFDFEQLRGFHFQVEGRDGGIPPRSATVTINLFVVDRNDNYPVILFPLPRNGSVP
                     VEIVPRSARTGHLVTKVVAEDADSGSNAWLSYHISRASDSSLFRISANIGELRTARLV
                     LPTDAVKQRVVVVVRDHGDPPLSSSVTLGVLLSNSVPQLLPDFEDVWEPGGQLSAQNL
                     YLVIALACISFLFLGCLLFFVCTKLHQSPGCCAQSCCRSTEDLRYGSKMVSNPCMTSA
                     TIDVTTVERLSQTYLYRASLGLGSDNNSLLLRGEYNAADLRNLATGVGLNLPISCIQI
                     RNRKGDHANVNAMPRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVS
                     SATPEPEAGEVSPPVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQ
                     EPTNSQIDKSDFITFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1385 a   1241 c   1223 g   1450 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5299;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4196 CAAGTCACAGAATTA 4182


RESULT  44
AF152304/c
LOCUS       AF152304     5431 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha C2 (PCDH-alpha-C2) mRNA, complete
            cds.
ACCESSION   AF152304
VERSION     AF152304.1  GI:5456893
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5431)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5431)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5431)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5431
                     /gene="PCDH-alpha-C2"
     CDS             1. .3024
                     /gene="PCDH-alpha-C2"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha C2"
                     /protein_id="AAD43698.1"
                     /db_xref="GI:5456894"
                     /translation="MEQAGTRPAATEHPRLRRPMPWLLLLPLLLLLLLLLPGPAASQL
                     RYSVPEEQAPGALVGNVARALGLELRRLGPGCLRINHLGAPSPRYLELDLTSGALFVN
                     ERIDREALCEQRPRCLLSLEVLAHNPVAVSAVEVEILDINDNSPRFPRPNYQLQVSES
                     VAPGARFHIESAQDPDVGANSVQTYELSPSEHFELDLKPLQENSKVLELVLRKGLDRE
                     QAALHHLVLTAVDGGIPARSGTAQISVRVLDTNDNSPAFDQSTYRVQLREDSPPGTLV
                     VKLNASDPDEGSNGELRYSLSSYTSDRERQLFSIDASTGEVRVIGGLDYEEASSYQIY
                     VQATDRGPVPMAGHCKVLVDIVDVNDNAPEVVLTDLYSPVPENATPNTIVAVLSVNDQ
                     DSGPNRKVSLGLEATLPFRLNGFGNSYTLVVSGPLDRERVAVYNITVTATDGGIPQLT
                     SLRTLKVEISDINDNPPSFLEDSYSIYIQENNLPGVLLCTVQATDPDEKENAEVTYSL
                     LEREIQGLPVTSYVSINSASGSLYAVNSFDYEKFREFFVTVEAQDKGSPPLSSTVTAN
                     VYVVDMNDHAPHILYPTSTNSSAAFEMVPRTAPAGYLVTKVIAMDSDSGQNAWLFYHL
                     AQTSDLDLFKVELHTGEIRTTRKMGDESGSTFNLTVVVRDNGEPSLSASVAITVAVVD
                     RVSKILPDTQRHVKSPRTYSEITLYLIIALSTVSFIFLLTIIILSIIKCYRYTAYGTA
                     CCGGFCGVRERSPAELYKQANNNIDARIPHGLKVQPHFIEVRGNGSLTKTYCYKACLT
                     AGSGSDTFMFYNTGAQTGPGPSGAQAAVTDSRNLTGQSGQNAGNLIILKNEAVSQNEP
                     RQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPP
                     VGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFIT
                     FGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1448 a   1318 c   1275 g   1390 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 89;  Length 5431;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4328 CAAGTCACAGAATTA 4314
```

```
RESULT  45
AB011138
LOCUS       AB011138     5562 bp    mRNA            PRI       10-APR-1998
DEFINITION  Homo sapiens mRNA for KIAA0566 protein, partial cds.
ACCESSION   AB011138
VERSION     AB011138.1  GI:3043655
KEYWORDS    KIAA0566 protein.
SOURCE      Homo sapiens male brain cDNA to mRNA, clone_lib:pBluescriptII SK
            plus clone:HH1910.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5562)
  AUTHORS   Ohara,O., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1998) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, DNA Technology; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
            Tel:+81-438-52-3913, Fax:+81-438-52-3914)
REFERENCE   2  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H.,
            Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes. IX.
            The complete sequences of 100 new cDNA clones from brain which can
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (1), 31-39 (1998)
  MEDLINE   98290545
FEATURES             Location/Qualifiers
     source          1. .5562
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HH1910"
                     /clone_lib="pBluescriptII SK plus"
                     /sex="male"
                     /tissue_type="brain"
     gene            1. .3494
                     /gene="KIAA0566"
     CDS             <1. .3494
                     /gene="KIAA0566"
                     /codon_start=3
                     /product="KIAA0566 protein"
                     /protein_id="BAA25492.1"
                     /db_xref="GI:3043656"
                     /translation="EKKSLFYVPKSDGSSLSPVTAAVYSFLTMIIVLQVLIPISLYVS
                     IEIVKACRVYFINQDMQLYDEETDSQLQCRALNITEDLGQIQYIFSDKTGTLTENKMV
                     FRRCTVSGVEYSHDANAQRLARYQEADSEEEEVVPRGGSVSQRGSIGSHQSVRVVHRT
                     QSTKSHRRTGSRAEAKRASMLSKHTAFSSPMEKDITPDPKLLEKVSECDKSLAVARHQ
                     EHLLAHLSPELSDVFDFFIALTICNTVVVTSPDQPRTKVRVRFELKSPVKTIEDFLRR
                     FTPSCLTSGCSSIGSLAANKSSHKLGSSFPSTPSSDGMLLRLEERLGQPTSAIASNGY
                     SSQADNWASELAQEQESERELRYEAESPDEAALVYAARAYNCVLVERLHDQVSVELPH
                     LGRLTFELLHTLGFDSVRKRMSVVIRHPLTDEINVYTKGADSVVMDLLQPCSSVDARG
                     RHQKKIRSKTQNYLNVYAAEGLRTLCIAKRVLSKEEYACWLQSHLEAESSLENSEELL
                     FQSAIRLETNLHLLGATGIEDRLQDGVPETISKLRQAGLQIWVLTGDKQETAVNIAYA
                     CKLLDHDEEVITLNATSQEACAALLDQCLCYVQSRGLQRAPEKTKGKVSMRFSSLCPP
                     STSTASGRRPSLVIDGRSLAYALEKNLEDKFLFLAKQCRSVLCCRSTPLQKSMVVKLV
                     RSKLKAMTLAIGDGANDVSMIQVADVGVGISGQEGMQAVMASDFAVPKFRYLERLLIL
                     HGHWCYSRLANMVLYFFYKNTMFVGLLFWFQFFCGFSASTMIDQWYLIFFNLLFSSLP
                     PLVTGVLDRDVPANVLLTNPQLYKSGQNMEEYRPRTFWFNMADAAFQSLVCFSIPYLA
                     YYDSNVDLFTWGTPIVTIALLTFLLHLGIETKTWTWLNWITCGFSVLLFFTVALIYNA
                     SCATCYPPSNPYWTMQALLGDPVFYLTCLMTPVAALLPRLFFRSLQGRVFPTQLQLAR
                     QLTRKSPRRCSAPKETFAQGRLPKDSGTEHSSGRTVKTSVPLSQPSWHTQQPVCSLEA
                     SGEPSTVDMSMPVREHTLLEGLSAPAPMSSAPGEAVLRSPGGCPEESKVRAASTGRVT
                     PLSSLFSLPTFSLLNWISSWSLVSRLGSVLQFSRTEQLADGQAGRGLPVQPHSGRSGL
                     QGPDHRLLIGASSRRSQ".
BASE COUNT     1347 a   1484 c   1472 g   1259 t
ORIGIN


  Query Match           60.0%;  Score 15;  DB 85;  Length 5562;
  Best Local Similarity  100.0%;  Pred. No. 32;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 tcgcaagtcacagaa 22
                   |||||||||||||||
Db    4666 TCGCAAGTCACAGAA 4680


Search completed: October  9, 2001, 15:43:40
Job time: 14606 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:39 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 24134

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	1041	20	AAX01303	E. coli biotin syn
2		25	100.0	1084	10	AAN91329	E.coli Bio B gene.
3		25	100.0	1121	7	AAN60496	Sequence encoding
4		25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	c 5	17	68.0	17	14	AAQ56307	BioB DNA primer.
6		15	60.0	561	22	AAF94235	Primer specific fo
7		15	60.0	1771	20	AAZ30602	Mouse integrin alp
	c 8	15	60.0	3520	22	AAF93909	Human cDNA encodin
	c 9	15	60.0	3520	22	AAF29361	Amyloid-beta prote
	c 10	15	60.0	6567	11	AAQ03324	Eimeria tenella ge
	c 11	14	56.0	1050	9	AAN80927	Lividans Exported
	c 12	14	56.0	1677	17	AAT31932	Retinoid X recepto
	c 13	14	56.0	1787	17	AAT31928	Retinoid X recepto
	c 14	14	56.0	1890	20	AAZ52940	Human prostate tum
	c 15	14	56.0	2233	19	AAV10120	Human retinoid rec
	c 16	14	56.0	2585	22	AAF77541	Caenorhabditis bri
	c 17	14	56.0	3083	12	AAQ11851	Glutamate receptor
	c 18	14	56.0	5773	21	AAC76014	Human ORFX ORF1569
19		14	56.0	32351	21	AAF21307	Human low adenosin
20		14	56.0	32351	21	AAA35185	Human adenosine re
21		14	56.0	40298	21	AAF21311	Human low adenosin
22		14	56.0	40298	21	AAA35189	Human adenosine re
23		14	56.0	1038602	20	AAZ01425	Complete genome se
24		13	52.0	87	16	AAT22576	Human gene signatu
	c 25	13	52.0	167	16	AAT20233	Human gene signatu
	c 26	13	52.0	187	21	AAC26213	Human secreted pro
	c 27	13	52.0	363	21	AAA50159	Antibody 5H7 heavy
	c 28	13	52.0	378	13	AAQ25667	Sequence of the an
	c 29	13	52.0	408	21	AAA74604	Murine 5B3 antibod
	c 30	13	52.0	415	21	AAC56783	Eucalyptus grandis
	c 31	13	52.0	476	21	AAC09392	Human secreted pro
	c 32	13	52.0	559	22	AAF93531	cDNA encoding SRT
33		13	52.0	637	21	AAA95830	Bovine metalloprot
	c 34	13	52.0	713	15	AAQ81957	CZP2(487-713). Ca
	c 35	13	52.0	777	18	AAT79355	Staphylococcus aur
	c 36	13	52.0	778	20	AAX32456	Partial nucleotide
	c 37	13	52.0	902	21	AAZ56756	Human transmembran
	c 38	13	52.0	930	20	AAX15335	cDNA encoding a le
	c 39	13	52.0	1002	21	AAA50164	DNA encoding 5H7 s
40		13	52.0	1010	21	AAC64778	Strawberry alcohol
41		13	52.0	1010	21	AAA27678	Strawberry alcohol
	c 42	13	52.0	1278	21	AAC36039	Arabidopsis thalia
	c 43	13	52.0	1317	21	AAA94840	Drosophila melanog
	c 44	13	52.0	1359	21	AAA72281	Drosophila odorant
	c 45	13	52.0	1404	21	AAZ59716	DNA encoding a mut

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     21 gacattgtcgcaagtcacagaatta 45


RESULT   2
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     44 gacattgtcgcaagtcacagaatta 68


RESULT   3
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     62 gacattgtcgcaagtcacagaatta 86


RESULT   4
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
```

```
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag=  a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag=  b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag=  c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag=  d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag=  e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag=  f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag=  g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag=  h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag=  i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag=  j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag=  k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag=  l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag=  m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag=  n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag=  o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag=  p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag=  q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag=  r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 gacattgtcgcaagtcacagaatta 161


RESULT   5
AAQ56307/c
ID   AAQ56307 standard; DNA; 17 BP.
XX
AC   AAQ56307;
XX
DT   13-APR-1994  (first entry)
XX
DE   BioB DNA primer.
XX
KW   BioA; BioB; promoter; biotin; operon; primer; ss.
XX
OS   Synthetic.
XX
PN   JP05219956-A.
```

```
XX
PD   31-AUG-1993.
XX
PF   14-SEP-1992;   92JP-0244792.
XX
PR   14-SEP-1992;   92JP-0244792.
XX
PA   (SHIS ) SHISEIDO CO LTD.
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
DR   WPI; 1993-308323/39.
XX
PT   DNA sequence of biotin operon - has base sequence of E. coli
PT   mutated by base pair(s) compared to wild type
XX
PS   Example 1; Fig 8; 11pp; Japanese.
XX
CC   A novel DNA sequence comprises the E.coli biotin operon (BO) in which
CC   the control region of BO or the region near the bioB initiation
CC   codon is mutated by at least one base pair compared to its
CC   wild type.  Two primers (AAQ56306-Q56307) are described in Example 1.
CC   A microorganism belonging to Escherichia genus, transformed by
CC   a recombinant plasmid carrying such DNA can be used for the prodn.
CC   of biotin-active substances.
XX
SQ   Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 14;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.34;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gtcgcaagtcacagaat 23
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1


RESULT   6
AAF94235
ID   AAF94235 standard; DNA; 561 BP.
XX
AC   AAF94235;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS   Synthetic.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 5; SEQ ID 669; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 561;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     339 caagtcacagaatta 353


RESULT   7
AAZ30602
ID   AAZ30602 standard; DNA; 1771 BP.
XX
AC   AAZ30602;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Mouse integrin alpha 4 coding sequence.
XX
KW   Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW   very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW   vascular endothelial cell; vascular endothelium; migration; inflammation;
KW   atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW   metastasis; circulatory system; autoimmune disease; Grave's disease;
KW   Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS   Mus sp.
XX
PN   US5968826-A.
XX
PD   19-OCT-1999.
XX
PF   05-OCT-1998;   98US-0166203.
XX
PR   05-OCT-1998;   98US-0166203.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM,  Condon TP;
XX
DR   WPI; 1999-590416/50.
DR   P-PSDB; AAY44154.
XX
PT   Antisense inhibition of integrin alpha4 expression useful for treating
PT   inflammatory diseases such as atherosclerosis, allergies, asthma and
```

PT arthritis -
XX
PS Example 12; Column 71-74; 40pp; English.
XX
CC This sequence represents the coding region of the mouse integrin
CC alpha4 gene. The invention relates to the generation of antisense
CC oligonucleotides targeted to the integrin alpha4 gene which are used
CC for inhibiting expression of the integrin alpha4 mRNA or protein.
CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
CC called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
CC surfaces of leucocytes and vascular endothelial cells and mediates the
CC adhesion of leucocytes to the vascular endothelium prior to migration
CC into the surrounding tissues. This migration is an essential step in
CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a
CC potential therapeutic target for treating inflammatory diseases and
CC the damaging effects of excessive inflammation. These disorders include
CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
CC metastasis (VLA-4 is involved in migration of the tumor cells through
CC the extracellular matrix into the circulatory system). VLA-4 is also
CC involved in a number of autoimmune diseases such as Grave's disease,
CC Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.
CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of
CC hemopoietic stem cells in bone-marrow and in allograft rejection.
XX
SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1771;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaatt 24
||||||||||||||||
Db 312 gcaagtcacagaatt 326

RESULT 8
AAF93909/c
ID AAF93909 standard; cDNA; 3520 BP.
XX
AC AAF93909;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0256.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
DR P-PSDB; AAB88482.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 331; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 3182 CAAGTCACAGAATTA 3168

RESULT 9
AAF29361/c
ID AAF29361 standard; DNA; 3520 BP.
XX
AC AAF29361;
XX
DT 20-APR-2001 (first entry)
XX
DE Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
XX
KW Human; amyloid-beta protein; agglutination regulatory factor;
KW Alzheimer's disease; ds.
XX
OS Homo sapiens.
XX
PN WO200104299-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-JP04515.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 18-OCT-1999; 99US-0159586.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;
PI Arakawa H, Morita M;
XX
DR WPI; 2001-138347/14.
DR P-PSDB; AAB49771.
XX
PT Polynucleotide encoding Amyloid-beta protein agglutination-controlling
PT factor, useful for inhibiting or promoting agglutination or
PT sedimentation of amyloid-beta protein and in diagnosis and screening
PT drugs for Alzheimer's disease -

```
XX
PS   Claim 1; Page 58-63; 72pp; Japanese.
XX
CC   This invention relates to polynucleotides AAF29357 - AAF29361 which
CC   encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
CC   agglutination of amyloid beta protein. The protein and polynucleotide
CC   sequences are useful in the diagnosis of Alzheimer's disease. They are
CC   also useful for screening drugs which are useful for treating Alzheimer's
CC   disease.
XX
SQ   Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT  10
AAQ03324/c
ID   AAQ03324 standard; DNA; 6567 BP.
XX
AC   AAQ03324;
XX
DT   02-AUG-1990  (first entry)
XX
DE   Eimeria tenella genomic DNA encoding antigen GX5401FL.
XX
KW   Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
XX
OS   Eimeria tenella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..6567
FT                   /*tag=  a
XX
PN   WO9000403-A.
XX
PD   25-JAN-1990.
XX
PF   05-JUL-1989;   89WO-US02918.
XX
PR   05-JUL-1988;   88US-0215162.
XX
PA   (GENE-) GENEX CORP.
XX
PI   Anderson DM,  McCandliss RJ,  Strausberg SL,  Strausberg RL;
XX
DR   WPI; 1990-051586/07.
DR   P-PSDB; AAR05222.
XX
PT   Cloned gene or fragment encoding antigenic protein -
PT   which binds with antibodies against avian coccidia, and
PT   transformed cells used in vaccine
XX
PS   Claim 10; Page 93 and Fig 14; 134pp; English.
XX
CC   An E. tenella genomic library was screened with radioactively labeled
CC   cDNA encoding the GX5401 antigen to identify clones encoding extensions
CC   of the GX5401 coding sequence. Several plaques reactive with the cDNA
CC   were identified. DNA from clone number 533 was sequenced. This analysis
CC   showed that the coding sequence in this clone overlaps with the GX5401
CC   coding sequence and extends that sequence toward the N-terminal coding
CC   region. Analysis of the new sequence together with the GX5401 antigen
CC   coding sequence reveals an open reading frame encoding an Eimeria
CC   protein of about 250,000 daltons. The protein carries several repeated
CC   peptide sequences and is rich in cysteine residues. The open reading
CC   frame encodes a potential signal sequence for protein secretion. The
CC   coding sequence for this antigenic protein is given in AAN93324. Also new
CC   are an expression vector contg. the cloned gene, host cells transformed
CC   with the vector, and AP encoded by the cloned gene. The transformed cells
CC   are used in a vaccine to immunise birds against avian coccidiosis. By
CC   labelling the peptides, they can be used as a type-specific probe. The
CC   AP may also be used in an assay to detect Ab against the coccidia. The
CC   Abs are used to identify transformed cells contg. the DNA.
XX
SQ   Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 11;  Length 6567;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db    3486 GACATTGTCGCAAGT 3472


RESULT  11
AAN80927/c
ID   AAN80927 standard; DNA; 1050 BP.
XX
AC   AAN80927;
XX
DT   15-OCT-1990  (first entry)
XX
DE   Lividans Exported Protein gene.
XX
KW   Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
KW   protease inhibitor; probe; ss.
XX
OS   Streptomyces lividans strain 1326.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     394..498
FT                   /*tag=  a
FT   mat_peptide     499.. 825
FT                   /*tag=  b
FT                   /product=LEP-10
XX
PN   WO8801278-A.
XX
PD   25-FEB-1988.
XX
PF   17-AUG-1987;   87WO-US02009.
XX
PR   18-AUG-1986;   86US-0897245.
XX
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PA   (BERK/) TR BERKA.
XX
PI   Berka TR,  Fornwald JA,  Gorniak JG,  Rosenberg M;
XX
DR   WPI; 1988-063996/09.
DR   P-PSDB; AAP80944.
XX
PT   Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
XX
PS   Disclosure; ; pp; English.
XX
CC   The sequence can be cloned and used to produce recombinant LEP-10.
CC   The DNA can be used as a probe for identifying further coding
CC   sequences for protease inhibitors; the protein is useful as an
CC   assay reagent.
CC   See also AAN80926.
XX
SQ   Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 9;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db     251 AAGTCACAGAATTA 238


RESULT  12
AAT31932/c
ID   AAT31932 standard; DNA; 1677 BP.
XX
AC   AAT31932;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-2 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-2; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1356
FT                   /*tag=  a
FT   polyA_signal    1672-1677
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99739.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 59-60; 90pp; English.
XX
CC   A cDNA clone (AAT31932) codes for retinoid X receptor interacting
CC   protein RIP14-2 (AAR99739), a previously undescribed orphan member
CC   of the nuclear receptor superfamily which may be involved in the
CC   complex retinoid response.  It was isolated from a mouse liver
CC   cDNA library using an RIP14 probe.  Different RIP14-2 clones had
CC   different 5' sequences (see also AAT31935-37).  Another isoform,
CC   RIP14-1, was also detected (see also AAT31928).  The cDNA can be
CC   used for the recombinant prodn. of RIP14-2 in transformed host
CC   cells.
XX
SQ   Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  13
AAT31928/c
ID   AAT31928 standard; DNA; 1787 BP.
XX
AC   AAT31928;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-1 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..1466
FT                   /*tag=  a
FT   polyA_signal    1782..1787
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99735.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 53-54; 90pp; English.
XX
CC   A full-length cDNA clone (AAT31928) codes for retinoid X receptor
CC   interacting protein RIP14-1 (AAR99735), a previously undescribed
CC   orphan member of the nuclear receptor superfamily which may be
CC   involved in the complex retinoid response.  It was isolated from
CC   a mouse liver cDNA library using an RIP14 probe.  Another isoform,
CC   RIP14-2, was also detected (see also AAT31932).  The cDNA can be used
CC   for the recombinant prodn. of RIP14-1 in transformed host cells.
XX
SQ   Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  14
AAZ52940/c
ID   AAZ52940 standard; cDNA; 1890 BP.
XX
AC   AAZ52940;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #83.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
```

```
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; AAY74061, AAY74062, AAY74063.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins -
XX
PS   Claim 2; Page 250; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1890;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1475 CAAGTCACAGAATT 1462


RESULT  15
AAV10120/c
ID   AAV10120 standard; cDNA; 2233 BP.
XX
AC   AAV10120;
XX
DT   29-MAY-1998  (first entry)
XX
DE   Human retinoid receptor RR1 cDNA.
XX
KW   Retinoid receptor; RR1; steroid receptor; agonist; antagonist; cancer;
KW   adrenal deficiency; skin disorder; inflammatory disorder;
KW   immune response regulator; autoimmune disease; therapeutic antibody; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             363..1781
FT                   /*tag=  a
FT                   /product= RR1
FT                   /note= "retinoid receptor"
XX
PN   US5728548-A.
XX
PD   17-MAR-1998.
XX
PF   29-JUN-1995;   95US-0496631.
XX
PR   29-JUN-1995;   95US-0496631.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Bowman M;
XX
DR   WPI; 1998-206567/18.
DR   P-PSDB; AAW40072.
XX
PT   Human retinoid receptor protein RR1 - useful for, e.g. drug
PT   screening, therapy and antibody production
XX
PS   Claim 8; Column 11-15; 13pp; English.
XX
CC   This sequence encodes a novel human steroid receptor, the retinoid
CC   receptor protein or RR1. This protein can be used in screening assays
CC   for steroid hormone receptor agonists and antagonists and in
CC   pharmaceutical compositions for treating adrenal deficiencies, e.g.
CC   Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
CC   inflammatory disorders, e.g. arthritis and HIV infections. The protein
CC   can also be used for regulating immune responses, e.g. as antitumour
CC   agents, vaccine adjuvants, organ rejection inhibitors or agents for
CC   treating autoimmune diseases. The protein can further be used to produce
CC   therapeutic antibodies.
XX
SQ   Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 19;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db    1055 TTGTCGCAAGTCAC 1042


RESULT  16
AAF77541/c
ID   AAF77541 standard; DNA; 2585 BP.
XX
AC   AAF77541;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Caenorhabditis briggsae mab-21 coding sequence.
XX
KW   Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
KW   chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
OS   Caenorhabditis briggsae.
XX
PN   US6197504-B1.
XX
PD   06-MAR-2001.
XX
PF   19-JAN-1998;   98US-0008697.
XX
PR   10-APR-1996;   96US-0011607.
PR   09-APR-1997;   97US-0835604.
XX
PA   (CHOW/) CHOW K L.
XX
PI   Chow KL;
XX
DR   WPI; 2001-256362/26.
DR   P-PSDB; AAB74419.
XX
PT   Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
PT   novel protein required for choice of alternate cell fates in sample, by
PT   contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
PS   Disclosure; column 41-44; 79pp; English.
XX
CC   The present invention describes a method of detecting mab-21 expression
CC   in a sample involving contacting total mRNA from the sample with a probe
CC   to the mab-21 gene. The sequences of the mab-21 coding sequence and
CC   protein from several species are also given. The human mab-21 gene is
CC   found on chromosome 13q13, a region which has been linked to Moebius
```

```
CC   syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC   are associated with aniridia and which is linked to peripheral neural
CC   tissue differentiation.
XX
SQ   Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 caagtcacagaatt 24
          ||||||||||||||
Db   1058 CAAGTCACAGAATT 1045


RESULT  17
AAQ11851/c
ID   AAQ11851 standard; DNA; 3083 BP.
XX
AC   AAQ11851;
XX
DT   31-JUL-1991  (first entry)
XX
DE   Glutamate receptor 3.
XX
KW   Glutamate receptor 3; probe; ligand; drug screening; ss.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             167..2830
FT                   /*tag=  a
FT                   /product= GR3
FT   sig_peptide     167..232
FT                   /*tag=  b
FT   mat_peptide     233..2830
FT                   /*tag=  c
XX
PN   WO9106648-A.
XX
PD   16-MAY-1991.
XX
PF   25-OCT-1990;   90WO-US06153.
XX
PR   27-OCT-1989;   89US-0428116.
XX
PA   (SALK ) SALK INST FOR BIOL STUD.
XX
PI   Heinemann SF,  Boulter JR,  Hollmann M,  Bettler B,  Jensen JE;
XX
DR   WPI; 1991-164197/22.
DR   P-PSDB; AAR11991.
XX
PT   Glutamate receptors - used to screen for functional ligands and
PT   identify and isolate further receptors
XX
PS   Disclosure; Fig 4; 109pp; English.
XX
CC   GluR3 cDNA was isolated from a rat forebrain cDNA using a
CC   low-stringency screening protocol and a radiolabelled fragment of the
CC   GluR1 cDNA as probe.  The cDNA is deposited (ATCC 68133).
CC   The gene and protein can be used in drug screening, to
CC   determine whether a substance is a functional ligand for the
CC   receptor by monitoring ion channel activity.
CC   See also AAQ11849-855.
XX
SQ   Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 12;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 cattgtcgcaagtc 16
          ||||||||||||||
Db   3034 CATTGTCGCAAGTC 3021


RESULT  18
AAC76014/c
ID   AAC76014 standard; cDNA; 5773 BP.
XX
AC   AAC76014;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB41805.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 2354-2357; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
```

```
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 5773;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaat 23
           ||||||||||||||
Db    4611 GCAAGTCACAGAAT 4598


RESULT  19
AAF21307
ID   AAF21307 standard; DNA; 32351 BP.
XX
AC   AAF21307;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2874.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1295-1303; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 32351;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db  10696 aagtcacagaatta 10709


RESULT  20
AAA35185
ID   AAA35185 standard; DNA; 32351 BP.
XX
AC   AAA35185;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1212-1219; 1343pp; English.
XX
```

```
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 21;  Length 32351;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 aagtcacagaatta 25
          ||||||||||||||
Db  10696 aagtcacagaatta 10709


RESULT  21
AAF21311
ID   AAF21311 standard; DNA; 40298 BP.
XX
AC   AAF21311;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2878.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1305-1315; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 21;  Length 40298;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 aagtcacagaatta 25
          ||||||||||||||
Db  11521 aagtcacagaatta 11534


RESULT  22
AAA35189
ID   AAA35189 standard; DNA; 40298 BP.
XX
AC   AAA35189;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
```

XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 1221-1231; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 40298;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aagtcacagaatta 25
||||||||||||||
Db 11521 aagtcacagaatta 11534

RESULT 23
AAZ01425
ID AAZ01425 standard; DNA; 1038602 BP.
XX
AC AAZ01425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal uretritis,
CC epidymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 56.0%; Score 14; DB 20; Length 1038602;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 attgtcgcaagtca 17
||||||||||||||
Db 736214 attgtcgcaagtca 736227

RESULT 24
AAT22576
ID AAT22576 standard; cDNA to mRNA; 87 BP.
XX
AC AAT22576;
XX
DT 01-OCT-1996 (first entry)
XX
DE Human gene signature HUMGS04192.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.

```
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1160; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;


  Query Match             52.0%;  Score 13;  DB 16;  Length 87;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db      16 aagtcacagaatt 28


RESULT  25
AAT20233/c
ID   AAT20233 standard; cDNA to mRNA; 167 BP.
XX
AC   AAT20233;
XX
DT   24-JUL-1996  (first entry)
XX
DE   Human gene signature HUMGS01380.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 592; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;


  Query Match             52.0%;  Score 13;  DB 16;  Length 167;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     118 AGTCACAGAATTA 106


RESULT  26
AAC26213/c
ID   AAC26213 standard; cDNA; 187 BP.
XX
AC   AAC26213;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30288.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30288; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
```

```
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;


  Query Match            52.0%;  Score 13;  DB 21;  Length 187;
  Best Local Similarity  100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 aagtcacagaatt 24
          |||||||||||||
Db    163 AAGTCACAGAATT 151


RESULT  27
AAA50159/c
ID   AAA50159 standard; cDNA; 363 BP.
XX
AC   AAA50159;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Antibody 5H7 heavy chain variable region cDNA.
XX
KW   5H7; monoclonal antibody; apoptosis; programmed cell death;
KW   major histocompatibility complex class I; MHC; tumour; therapy;
KW   graft-mediated immunoprotection; single chain antibody; scFv; ss.
XX
OS   Mammalia.
XX
PN   WO200047713-A2.
XX
PD   17-AUG-2000.
XX
PF   08-FEB-2000; 2000WO-US03234.
XX
PR   09-FEB-1999;   99US-0119238.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Woodle ES,  Van Seventer JM,  Kulkarni S,  Kranz D,  Holman P;
XX
DR   WPI; 2000-558210/51.
DR   P-PSDB; AAY95776.
XX
PT   New polynucleotides encoding single chain light and heavy variable
PT   regions of a 5H7 antibody, useful in methods for conferring programmed
PT   cell death and graft-mediated immune protection of cell, tissue and
PT   organs -
XX
PS   Claim 2; Page 36; 48pp; English.
XX
CC   The present sequence is that of cDNA coding for the heavy chain
CC   variable region (VH) (see AAY95776) of 5H7, an anti-human class I
CC   major histocompatibility complex (MHC) monoclonal antibody (MAb)
CC   which recognises a monomorphic determinant of the alpha-3 domain.
CC   MAb 5H7 has been shown to induce programmed cell death (PCD) in
CC   lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
CC   splenocytes transgenic for class I HLA-B27.  Polynucleotides
CC   encoding 5H7 VL and VH sequences were used in the construction of
CC   single chain variable immunoglobulin domains (scFv) of the 5H7
CC   antibody (see AAY95781).  These can be used to confer PCD properties
CC   to cells, especially T and B lymphocyte tumour cells.  Recombinant
CC   DNA molecules encoding 5H7 scFv are also used in methods for
CC   conferring graft-mediated immune protection of cell, tissue and
CC   transplanted organs.
XX
SQ   Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;


  Query Match            52.0%;  Score 13;  DB 21;  Length 363;
  Best Local Similarity  100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 aagtcacagaatt 24
          |||||||||||||
Db    262 AAGTCACAGAATT 250


RESULT  28
AAQ25667/c
ID   AAQ25667 standard; cDNA; 378 BP.
XX
AC   AAQ25667;
XX
DT   28-DEC-1992  (first entry)
XX
DE   Sequence of the anti-urokinase antibody kappa variable region (VK)
DE   cDNA.
XX
KW   Chimeric monoclonal antibody; anti-urokinase antibody; PCR;
KW   antithrombotic agent; myocardial infarction therapy; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..9
FT                   /*tag=  a
FT   mat_peptide     10..378
FT                   /*tag=  b
FT   misc_feature    1..27
FT                   /*tag=  c
FT                   /label= VH01 primer
FT   misc_feature    18..32
FT                   /*tag=  d
FT                   /label= 5'VH1 primer
FT   misc_feature    340..371
FT                   /*tag=  e
FT                   /label= 3'mVH primer
FT   misc_feature    356..371
FT                   /*tag=  f
FT                   /label= JH01 primer
XX
PN   EP491351-A.
XX
PD   24-JUN-1992.
XX
PF   17-DEC-1991;   91EP-0121591.
XX
PR   18-DEC-1990;   90JP-0413829.
PR   11-NOV-1991;   91JP-0294464.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Iwasa S,  Taka H,  Watanabe T,  Tada H;
XX
DR   WPI; 1992-209528/26.
DR   P-PSDB; AAR24722.
XX
PT   Chimeric monoclonal antibodies - contain anti-human fibrin
PT   antibody light and heavy chain variable and constant for treating
PT   thrombotic conditions e.g. myocardial infarction
XX
PS   Example; Figure 15; 87pp; English.
XX
CC   Poly(A)+ RNA was prepd. from mouse anti-urokinase
CC   antibody-producing hybridoma UK1-3 cells. Using this poly(A) RNA as
CC   a template, an anti-urokinase antibody VK cDNA was amplified with
```

CC the mC-kappa primer as primer for first strand sysnthesis, and the
CC 3'mV-kappa and 5'mV-kappa primers as primers for the PCR. The
CC amplified fragment was restriction digested and ligated into a
CC restriction fragment of pTB1423 to give an anti-urokinase antibody
CC VK cDNA contg. plasmid pTB1456. The sequence of this plasmid is
CC given in AAQ25667. The cDNA is a functional VK gene. The sequence of
CC the primer (mC-gamma-1) is given in AAQ25689.
XX
SQ Sequence 378 BP; 92 A; 100 C; 91 G; 95 T; 0 other;

Query Match 52.0%; Score 13; DB 13; Length 378;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aagtcacagaatt 24
||||||||||||
Db 271 AAGTCACAGAATT 259

RESULT 29
AAA74604/c
ID AAA74604 standard; cDNA; 408 BP.
XX
AC AAA74604;
XX
DT 08-JAN-2001 (first entry)
XX
DE Murine 5B3 antibody heavy chain variable region cDNA.
XX
KW Mouse; 5B3 antibody; IgG1; piezoelectric immunoassay;
KW small molecule explosive detection; 2,4,6-trinitrotoluene; TNT; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..408
FT /*tag= a
FT /partial
FT /product= "5B3 antibody heavy chain variable region"
XX
PN WO200043774-A2.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-IL00048.
XX
PR 25-JAN-1999; 99IL-0128212.
XX
PA (YISS) YISSUM RES & DEV CO.
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Willner I, Eshhar Z;
XX
DR WPI; 2000-524259/47.
DR P-PSDB; AAB15672.
XX
PT Apparatus for detecting small molecules, especially explosives
PT comprises a piezoelectric crystal -
XX
PS Disclosure; Fig 3A; 90pp; English.
XX
CC The present sequence encodes the heavy chain variable region of 5B3
CC antibody. It was obtained from total RNA extracted from a 5B3 hybridoma
CC by RT-PCR. 5B3 is an IgG1 antibody derived from a TNP-KLH immunised
CC mouse. Its binding to TNP or DNS antigen can be blocked by very low
CC amounts of TNT and it can therefore be used in a method for detecting
CC small assayed explosive molecules. Molecules are detected using a
CC piezoelectric sensor. Piezoelectric immunoassaying in liquid phase
CC allows stationary and flow analysis of an aqueous sample. The method
CC is sufficiently sensitive for detection of low molecular weight
CC molecules.
XX
SQ Sequence 408 BP; 90 A; 108 C; 99 G; 111 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 408;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aagtcacagaatt 24
||||||||||||
Db 316 AAGTCACAGAATT 304

RESULT 30
AAC56783/c
ID AAC56783 standard; DNA; 415 BP.
XX
AC AAC56783;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #654.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Pages 513-514; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 415;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      4 attgtcgcaagtc 16
          |||||||||||||
Db    121 ATTGTCGCAAGTC 109


RESULT  31
AAC09392/c
ID   AAC09392 standard; cDNA; 476 BP.
XX
AC   AAC09392;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 13467.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 13467; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 476 BP; 109 A; 72 C; 90 G; 200 T; 5 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaat 23
           |||||||||||||
Db     263 CAAGTCACAGAAT 251


RESULT  32
AAF93531/c
ID   AAF93531 standard; cDNA; 559 BP.
XX
AC   AAF93531;
XX
DT   21-MAY-2001  (first entry)
XX
DE   cDNA encoding SRT protein isolated from testis tissue SEQ ID 352.
XX
KW   Human; SRT; gene therapy; gene mapping; tissue typing; ss.
XX
OS   Homo sapiens.
XX
PN   WO200107611-A2.
XX
PD   01-FEB-2001.
XX
PF   21-JUL-2000; 2000WO-US20006.
XX
PR   26-JUL-1999;   99US-0145701.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Goddard A,  Wood WI;
XX
DR   WPI; 2001-112729/12.
XX
PT   New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT   for production of recombinant SRT polypeptides, gene mapping,
PT   diagnosing genetic disorders and for gene therapy -
XX
PS   Claim 2; Fig 352; 663pp; English.
XX
CC   Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC   human SRT proteins. The cDNA sequences are isolated from various
CC   different human tissue cDNA libraries. The invention relates to a method
CC   for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC   encoding SRT, a host cell transformed with the vector, an isolated SRT
CC   polypeptide, and an antibody which binds to SRT. The polynucleotide
CC   sequence can be used in gene therapy and is useful in the recombinant
CC   production of SRT polypeptides, as a hybridisation probe to screen
CC   libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC   map the gene encoding the SRT polypeptides and analysing genetic
CC   disorders, tissue typing and disease tissue detection. The SRT
CC   polynucleotide sequences can be used in polymerase chain reaction,
CC   screening for new therapeutic molecules and generation of antisense RNA
CC   and DNA.
XX
SQ   Sequence 559 BP; 127 A; 127 C; 130 G; 171 T; 4 other;


  Query Match             52.0%;  Score 13;  DB 22;  Length 559;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaat 23
           |||||||||||||
Db      78 CAAGTCACAGAAT 66


RESULT  33
AAA95830
ID   AAA95830 standard; cDNA; 637 BP.
XX
AC   AAA95830;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Bovine metalloproteinase KIAA0688 cDNA.
XX
KW   Bovine; KIAA0688; metalloproteinase; ADAM;
KW   a disintegrin and metalloproteinase domain; thrombospondin domain;
KW   vaccine; nootropic; neuroprotective; antiparkinsonian;
KW   cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW   Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW   autoimmune disease; brain tumour; brain injury; ss.
XX
OS   Bos taurus.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             2..370
FT                   /*tag=  a
FT                   /partial
FT                   /product= "KIAA0688"
XX
PN   WO200053774-A2.
XX
PD   14-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US06237.
XX
PR   08-MAR-1999;   99US-0264585.
XX
PA   (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PI   Kelner GS,  Clark M,  Maki RA;
XX
DR   WPI; 2000-594326/56.
DR   P-PSDB; AAB21260.
XX
PT   Polynucleotide encoding novel members of a disintegrin,
PT   metalloproteinase and thrombospondin domain protein family used to
PT   prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
PS   Example 3; Fig 20; 129pp; English.
XX
CC   The present sequence encodes bovine metalloproteinase KIAA0688, which is
CC   a member of the ADAMTS family of proteins. Members of the ADAMTS family
CC   contain a metalloproteinase, integrin and thrombospondin domain. ADAMTS
CC   polypeptides are useful for the manufacture of medicaments for treating
CC   conditions associated with neuroinflammation and/or neurodegeneration,
CC   such as Alzheimer's disease, Parkinson's disease and stroke. They are
CC   also useful for treating conditions associated with cell proliferation,
CC   cell migration, inflammation and/or angiogenesis, such as cancer,
CC   arthritis and autoimmune diseases. They can be used to treat patients
CC   afflicted with an invasive tumour, a brain tumour or brain injury.
XX
SQ   Sequence 637 BP; 107 A; 227 C; 196 G; 107 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 637;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 cgcaagtcacaga 21
           |||||||||||||
Db     314 cgcaagtcacaga 326


RESULT  34
AAQ81957/c
ID   AAQ81957 standard; cDNA to mDNA; 713 BP.
XX
AC   AAQ81957;
XX
DT   15-MAR-1995  (first entry)
XX
DE   CZP2(487-713).
XX
KW   Canine; dog; zona pellucida; ZP; CZP2; contraceptive; vaccine;
KW   antigen; ss.
XX
OS   Canis familiaris.
XX
PN   JP06189766-A.
XX
PD   12-JUL-1994.
XX
PF   25-DEC-1992;   92JP-0359265.
XX
PR   25-DEC-1992;   92JP-0359265.
XX
PA   (TOFU ) TONEN CORP.
XX
DR   WPI; 1994-259553/32.
XX
PT   New DNA sequence encoding canine zona pellucida CZP2 - useful for
PT   the prodn. of a canine contraceptive vaccine antigen
XX
PS   Disclosure; Page 6; 10pp; Japanese.
XX
CC   The CZP2 DNA (AAQ70072) was prepd. by the cloning of CZP2(75-520) -
CC   AAQ81700 using the primers given in AAQ70073-74, CZP2(1-65) - AAQ81804
CC   using the primers given in AAQ70082-83, CZP2(42-103) - AAQ81803 using
CC   the primers given in AAQ70079-81 and CZP2(487-713) - AAQ81957 using the
CC   primers given in AAQ70075-78.
XX
SQ   Sequence 713 BP; 177 A; 189 C; 162 G; 185 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 15;  Length 713;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db     429 GCAAGTCACAGAA 417


RESULT  35
AAT79355/c
ID   AAT79355 standard; DNA; 777 BP.
XX
AC   AAT79355;
XX
DT   23-MAR-1998  (first entry)
XX
DE   Staphylococcus aureus downstream junction sequence LO2C4.
XX
KW   Staphylococcus aureus; junction; upstream; downstream; mec domain;
KW   methicillin resistance; coagulase negative; ds.
XX
OS   Staphylococcus aureus.
XX
PN   WO9731125-A2.
XX
PD   28-AUG-1997.
XX
PF   21-FEB-1997;   97WO-JP00487.
XX
PR   23-FEB-1996;   96JP-0060373.
XX
PA   (HIRA/) HIRAMATSU K.
PA   (KAIN-) KAINOS LAB INC.
XX
PI   Awaya A,  Hayashi T,  Ohno H;
XX
DR   WPI; 1997-435174/40.
XX
PT   Diagnosing methicillin resistant Staphylococcus aureus or
PT   coagulase-negative Staphylococcus - by detecting junction sequences
PT   at upstream and downstream points where methicillin resistance
PT   domain adjoins rest of chromosome
XX
PS   Claim 5; Fig 12; 71pp; Japanese.
XX
CC   This LO2C4 is the downstream junction sequence of a Staphylococcus
CC   aureus DNA. This is used in a method for diagnosing methicillin resistant
CC   Staphylococcus aureus (MRSA) or methicillin resistant coagulase-negative
CC   Staphylococcus (MRC-NS). The method comprises extracting nucleic acid
CC   molecules from a Staphylococci containing sample, and detecting the
CC   junction sequences at the upstream and downstream points where the
CC   methicillin resistance (mec) domain adjoins the rest of the chromosome.
CC   The sequences are subjected to NASBA-dot hybridisation. The method can be
```

```
CC   used for the detection of MRSA and MRC-NS in biological samples such as
CC   blood, urine and sputum, and their discrimination from methicillin-
CC   sensitive strains.
XX
SQ   Sequence 777 BP; 249 A; 130 C; 91 G; 307 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 18;  Length 777;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     636 AGTCACAGAATTA 624


RESULT  36
AAX32456/c
ID   AAX32456 standard; DNA; 778 BP.
XX
AC   AAX32456;
XX
DT   22-JUN-1999  (first entry)
XX
DE   Partial nucleotide sequence of NCTC10442.
XX
KW   Methicillin-resistant coagulase-negative Staphylococci; MRC-NS; LCR;
KW   mecA; hybridisation; MSC-NS; MRSA; methicillin-resistant S. aureus;
KW   PCR; NASBA; ss.
XX
OS   Staphylococcus sp.
XX
PN   JP11056371-A.
XX
PD   02-MAR-1999.
XX
PF   22-AUG-1997;   97JP-0242070.
XX
PR   22-AUG-1997;   97JP-0242070.
XX
PA   (HIRA/) HIRAMATSU K.
PA   (ITOK/) ITO K.
PA   (KAIN-) KAINOSU KK.
XX
DR   WPI; 1999-222384/19.
XX
PT   Identification of methicillin-resistant Staphylococcus aureus (MRSA)
PT   or methicillin-resistant coagulase-negative Staphylococci (MRC-NS) -
PT   useful for rapid analysis of biological samples
XX
PS   Examples; Fig 15-16; 45pp; Japanese.
XX
CC   The invention relates to a method for the identification of methicillin-
CC   resistant coagulase-negative Staphylococci (MRC-NS) in which part of
CC   mecDNA which is a foreign inserted DNA present on the chromosome of MRC-
CC   NS and carrying mecA gene and part of the base sequence of the chromosome
CC   DNA surrounding the inserted DNA are used together to react with a sample
CC   specimen. The method identifies MRC-NS by NASBA, PCR, LCR or
CC   hybridisation by utilising a fact that the base sequence of a chromosome
CC   DNA surrounding the inserted site of mecDNA in the chromosome of
CC   methicillin sensitive coagulase negative Staphylococci (MSC-NS) is used
CC   for a reaction with a sample and when mecDNA is inserted it comes
CC   negative. The method can be used for especially identifying methicillin-
CC   resistant Staphylococcus aureus (MRSA). The method can detect MRSA from a
CC   specimen such as urine, blood, sputum, etc. in 30 minutes to 1 hour.
XX
SQ   Sequence 778 BP; 249 A; 130 C; 91 G; 308 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 778;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     637 AGTCACAGAATTA 625


RESULT  37
AAZ56756/c
ID   AAZ56756 standard; cDNA; 902 BP.
XX
AC   AAZ56756;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human transmembrane protein HTMPN-59 encoding cDNA.
XX
KW   Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW   antiproliferative; neuroprotective; immune disorder;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   gastrointestinal disorder; developmental disorder;
KW   cell proliferative disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9961471-A2.
XX
PD   02-DEC-1999.
XX
PF   28-MAY-1999;   99WO-US11904.
XX
PR   29-MAY-1998;   98US-0087260.
PR   02-JUL-1998;   98US-0091674.
PR   02-OCT-1998;   98US-0102954.
PR   24-NOV-1998;   98US-0109869.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Hillman JL,  Yue H,  Guegler KJ,  Corley NC;
PI   Bandman O,  Patterson C,  Gorgone GA,  Kaser MR,  Baughn MR;
PI   Au-Young J;
XX
DR   WPI; 2000-072605/06.
DR   P-PSDB; AAY57935.
XX
PT   Proteins, polynucleotides, vectors, host cells and antibodies used to
PT   diagnose, treat or prevent immune, reproductive, smooth muscle,
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders -
XX
PS   Claim 9; Page 215; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neurological,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 902;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     425 AGTCACAGAATTA 413
```

```
RESULT  38
AAX15335/c
ID   AAX15335 standard; cDNA; 930 BP.
XX
AC   AAX15335;
XX
DT   20-MAY-1999  (first entry)
XX
DE   cDNA encoding a lectin recognising glucose.
XX
KW   Lectin protein; glucose; maboya; fungal disease; Candidosis;
KW   Aspergillosis; Cryptococcosis; mucormycosis; trichosporonosis; ss.
XX
OS   Halocynthia roretzi.
XX
FH   Key             Location/Qualifiers
FT   CDS             52..726
FT                   /*tag=  a
XX
PN   JP11032759-A.
XX
PD   09-FEB-1999.
XX
PF   17-JUL-1997;   97JP-0208580.
XX
PR   17-JUL-1997;   97JP-0208580.
XX
PA   (FUJI/) FUJITA T.
XX
DR   WPI; 1999-183820/16.
DR   P-PSDB; AAW97115.
XX
PT   A new lectin protein and DNA - useful for prevention and treatment
PT   of fungal diseases such as Candidosis, Aspergillosis,
PT   Cryptococcosis, mucormycosis and trichosporonosis
XX
PS   Disclosure; Fig 1; 9pp; Japanese.
XX
CC   The present sequence encodes a lectin protein that recognises
CC   glucose. The protein is isolated from maboya body fluid. The protein
CC   is useful in agents for prevention and treatment of fungal diseases
CC   such as Candidosis, Aspergillosis, Cryptococcosis, mucormycosis and
CC   trichosporonosis.
XX
SQ   Sequence 930 BP; 320 A; 177 C; 203 G; 230 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 930;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 attgtcgcaagtc 16
          |||||||||||||
Db    437 ATTGTCGCAAGTC 425


RESULT  39
AAA50164/c
ID   AAA50164 standard; DNA; 1002 BP.
XX
AC   AAA50164;
XX
DT   07-NOV-2000  (first entry)
XX
DE   DNA encoding 5H7 single chain antibody (scFv).
XX
KW   5H7; monoclonal antibody; apoptosis; programmed cell death;
KW   major histocompatibility complex class I; MHC; tumour; therapy;
KW   graft-mediated immunoprotection; scFv; single chain antibody; ss.
XX
OS   Chimeric - mammalia.
OS   Chimeric - synthetic.
XX
PN   WO200047713-A2.
XX
PD   17-AUG-2000.
XX
PF   08-FEB-2000; 2000WO-US03234.
XX
PR   09-FEB-1999;   99US-0119238.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Woodle ES,  Van Seventer JM,  Kulkarni S,  Kranz D,  Holman P;
XX
DR   WPI; 2000-558210/51.
DR   P-PSDB; AAY95781.
XX
PT   New polynucleotides encoding single chain light and heavy variable
PT   regions of a 5H7 antibody, useful in methods for conferring programmed
PT   cell death and graft-mediated immune protection of cell, tissue and
PT   organs -
XX
PS   Disclosure; Page 41-42; 48pp; English.
XX
CC   The present sequence is that of a recombinant DNA molecule encoding
CC   a single chain antibody (scFv).  The scFv is based on 5H7, an
CC   anti-human class I major histocompatibility complex (MHC) monoclonal
CC   antibody (MAb) that confers programmed cell death (PCD) inducing
CC   properties to cells.  The scFv is composed of (from the N-terminus)
CC   a leader peptide, the light chain variable region of MAb 5H7, a
CC   linker peptide, the heavy chain variable region of MAb 5H7, a
CC   second linker peptide, and glycosylphosphotidylinostitol.  The
CC   recombinant DNA molecule can be placed into a variety of expression
CC   systems and hosts for production of the scFv.  A method for
CC   conferring PCD inducing properties to cells involves transforming a
CC   cell with the recombinant DNA molecule.  The method provides for PCD
CC   signalling in B and T lymphocyte tumour cells.  In a related
CC   method, graft-mediated immune protection is conferred to a cell,
CC   tissue or organ (especially an organ for transplant) by
CC   transformation with the recombinant DNA molecule.
XX
SQ   Sequence 1002 BP; 258 A; 258 C; 237 G; 249 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 1002;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    733 AAGTCACAGAATT 721


RESULT  40
AAC64778
ID   AAC64778 standard; cDNA; 1010 BP.
XX
AC   AAC64778;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Strawberry alcohol dehydrogenase encoding cDNA SEQ ID NO:13A.
XX
KW   Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;
KW   aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;
KW   pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;
KW   alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;
KW   food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;
KW   confectionery; flavouring; oral medication; vitamin; aroma; beverage;
KW   alcohol; scent; fragrance; perfume; cosmetic; suspension aid;
KW   aluminium salt; anti-perspirant; pharmaceutical; cleaning product;
KW   insect pheromone; dye carrier; solvent; insect repellent; miticide;
KW   scabicide; plasticiser; deodorant; ss.
XX
OS   Fragaria x ananassa.
```

```
XX
PN   WO200032789-A1.
XX
PD   08-JUN-2000.
XX
PF   02-DEC-1999;   99WO-NL00737.
XX
PR   02-DEC-1998;   98EP-0204018.
PR   12-MAR-1999;   99EP-0200739.
XX
PA   (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.
XX
PI   Aharoni A,  Luecker J,  Verhoeven HA,  Van Tunen AJ,  O'Connell AP;
XX
DR   WPI; 2000-412335/35.
DR   P-PSDB; AAB36451.
XX
PT   A new DNA sequence encoding a polypeptide with alcohol acyl transferase
PT   activity for producing and regulating aromatic and/or aliphatic ester
PT   formation in microorganisms, plant cells or plants -
XX
PS   Claim 37; Page 100-101; 163pp; English.
XX
CC   The present invention describes nucleotide sequences with thiolase,
CC   alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,
CC   aminotransferase and esterase activities, which are involved in the
CC   biosynthetic pathway for aliphatic and/or aromatic ester production in
CC   fruit. The nucleotide sequences can be inserted into the genome of a
CC   fruit-producing plant to regulate aliphatic and/or aromatic ester
CC   formation. Aromatic and/or aliphatic esters in microorganisms, plant
CC   cells or plants are produced by inserting thiolase, alcohol acyl
CC   transferase, alcohol dehydrogenase, pyruvate decarboxylase,
CC   aminotransferase and esterase nucleotide sequences into the genome and
CC   feeding the microorganism or plant with alcohol, aldehydes, alpha-keto
CC   acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and
CC   their proteins can be used in the processed food industry as food
CC   additives to enhance the flavour of syrups, ice-creams, frozen desserts,
CC   yoghurts and confectionery. They are used: as flavouring agents for oral
CC   medications and vitamins; provide flavour and aroma in beverages,
CC   including alcohol; enhance or reduce fruit flavour, aroma, fragrance or
CC   scent; enhance the flavour or aroma of natural, synthetic or artificial
CC   products; for the production of novel combinations of artificial flavour
CC   substances; as antibacterial or anti-fungal agents; as fragrance or
CC   perfumes in cosmetics, creams, sun-protectant products, hair
CC   conditioners, lengthening agents and fixatives in perfumes, suspension
CC   aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning
CC   products, personal care products and animal care products; as
CC   disinfectant additives; as degreasing solvents for electronics; as
CC   insect pheromones; and as dye carriers, solvents, insect repellents,
CC   miticides, scabicides, plasticisers and deodorants. The present sequence
CC   encodes the specifically claimed strawberry alcohol dehydrogenase.
XX
SQ   Sequence 1010 BP; 333 A; 171 C; 237 G; 269 T; 0 other;

  Query Match             52.0%;  Score 13;  DB 21;  Length 1010;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db     385 gcaagtcacagaa 397


RESULT  41
AAA27678
ID   AAA27678 standard; cDNA; 1010 BP.
XX
AC   AAA27678;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Strawberry alcohol dehydrogenase partial cDNA clone SLG144.
XX
KW   Strawberry; alcohol dehydrogenase; fruit; ripening; ester;
KW   flavour; aroma; transgenic plant; ss.
XX
OS   Fragaria ananassa.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..853
FT                   /*tag=  a
FT                   /partial
XX
PN   EP1006190-A1.
XX
PD   07-JUN-2000.
XX
PF   02-DEC-1998;   98EP-0204018.
XX
PR   02-DEC-1998;   98EP-0204018.
XX
PA   (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.
XX
PI   Verhoeven HA,  van Tunen AJ,  Aharoni A,  Luecker J,  O'Connell AP;
XX
DR   WPI; 2000-378264/33.
DR   P-PSDB; AAY79668.
XX
PT   New polynucleotides encoding enzymes from the biosynthetic pathway for
PT   aromatic and/or aliphatic ester production in fruit used to modify
PT   plant flavours -
XX
PS   Example 3; Page 93-94; 116pp; English.
XX
CC   The present sequence is that of a partial cDNA clone for strawberry
CC   cv. Elsanta alcohol dehydrogenase SLG144 (see AAY79668), an enzyme
CC   that shows upregulated expression in ripening fruit.  The clone was
CC   isolated from a cDNA library by expression analysis using Northern
CC   blotting.  The invention relates to DNA sequences (see AAA27666-78)
CC   encoding enzymes (see AAY29656-68) involved in the metabolic pathway
CC   leading to the formation of aliphatic and/or aromatic esters in
CC   ripening fruit.  The enzymes have alcohol acyl transferase,
CC   alcohol dehydrogenase, pyruvate decarboxylase, thiolase or
CC   aminotransferase activity.  Expression vectors comprising the DNA
CC   sequences may be used to regulate ester formation in fruit.
CC   Genetically modified plants, plant cells and microorganisms can be
CC   used to produce esters.  The DNA sequences, polypeptides and
CC   antibodies are also used to screen fruit: for volatile ester
CC   compounds; for quality such as flavour, fragrance, aroma, scent,
CC   texture or shape; to distinguish between cultivars and varieties;
CC   and to monitor harvest time, post-harvest quality, shelf-life,
CC   timing of pesticide application, and resistance capacity based on
CC   volatile ester profiles.
XX
SQ   Sequence 1010 BP; 333 A; 171 C; 237 G; 269 T; 0 other;

  Query Match             52.0%;  Score 13;  DB 21;  Length 1010;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db     385 gcaagtcacagaa 397


RESULT  42
AAC36039/c
ID   AAC36039 standard; DNA; 1278 BP.
XX
AC   AAC36039;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 12322.
```

XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             52.0%;  Score 13;  DB 21;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cattgtcgcaagt 15
           |||||||||||||
Db      78 CATTGTCGCAAGT 66


RESULT  43
AAA94840/c
ID   AAA94840 standard; DNA; 1317 BP.
XX
AC   AAA94840;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Drosophila melanogaster odorant receptor DOR44 coding sequence.
XX
KW   Odorant receptor; fruit fly; DOR44; odour recognition; pest control; ds.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1317
FT                   /*tag=  a
FT                   /product= "DOR44"
FT                   /partial
XX
PN   WO200050566-A2.
XX
PD   31-AUG-2000.
XX
PF   25-FEB-2000; 2000WO-US04995.
XX
PR   25-FEB-1999;   99US-0257706.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Vosshall LB,  Amrein HO,  Axel R;
XX
DR   WPI; 2000-572081/53.
DR   P-PSDB; AAB26441.
XX
PT   Novel nucleic acid encoding an insect odorant receptor, for identifying
PT   modulator compounds that are useful in controlling pest population  -
XX
PS   Disclosure; Page 70-71; 176pp; English.
XX
CC   The present sequence is the coding sequence for the previously identified
CC   Drosophila melanogaster odorant receptor DOR44. The odorant genes and
CC   proteins, such as those provided by the invention, are useful as they aid
CC   in the study of the olfactory organ in mammals, as well as aiding the
CC   understanding of the link between odour recognition and behaviour in
CC   insects. They also enable the identification of compounds capable of
CC   activating and inhibiting the receptors, allow the control of pest
CC   populations via the use of alarm odour ligands and via the use of ligands
CC   which interfere with the interaction between odorant ligands and
CC   receptors associated with fertility.
XX
SQ   Sequence 1317 BP; 299 A; 317 C; 327 G; 374 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 1317;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaat 23
           |||||||||||||
Db     162 CAAGTCACAGAAT 150


RESULT  44
AAA72281/c
ID   AAA72281 standard; cDNA; 1359 BP.
XX
AC   AAA72281;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Drosophila odorant receptor DORLU 25.1 cDNA.
XX
KW   Odorant receptor; Drosophila; olfactory receptor;
KW   G protein-coupled receptor; GPCR superfamily; transgenic insect;
KW   insect behaviour modification; pest control; pollinator attraction;
KW   biosensor; odour detection; odour identification; apiculture; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200043410-A2.
XX
PD   27-JUL-2000.
XX
PF   25-JAN-2000; 2000WO-US01823.
XX
PR   25-JAN-1999;   99US-0117132.
XX
PA   (UYYA ) UNIV YALE.
```

XX
PI Carlson JR, Kim J, Clyne PJ, Warr CG;
XX
DR WPI; 2000-543246/49.
DR P-PSDB; AAB20946.
XX
PT New nucleic acid encoding a Drosophila olfactory receptor, useful for
PT identifying modulating agents -
XX
PS Claim 3; Page 271-273; 303pp; English.
XX
CC Sequences AAA72236-A72284 represent cDNAs encoding Drosophila
CC melanogaster odorant receptors (AAB20901-B20949). These proteins function
CC as olfactory receptors, and are thought to be members of the G
CC protein-coupled receptor (GPCR) superfamily, which is characterised by
CC the presence of 7 transmembrane helices. Nucleic acids encoding the
CC Drosophila odorant receptors may be used to generate expression
CC constructs, host cells containing such constructs, and transgenic
CC insects. Cells which express the odorant receptor genes may be used in
CC methods to identify agents which modulate expression of these genes, and
CC in methods to identify receptor binding partners. The Drosophila odorant
CC receptor nucleic acids may also be used to identify corresponding genes
CC in other insects, such as those which damage crops or transmit disease.
CC The odorant receptor proteins may be used to identify agents which
CC modulate their activity, to identify binding partners, as antigens to
CC raise antibodies, and in methods to modify insect behaviour. The proteins
CC may be also be used in methods of behaviour modification. Such methods
CC may be used to study or modify insect behaviour in response to odorants
CC such as pheromones. Modification of insect behaviour has a wide range of
CC applications, such as in pest control (e.g., by disrupting the feeding
CC or mating behaviours of pest species), or for enhancing plant
CC pollination (by attracting pollinator species). Odorant receptor proteins
CC and/or nucleotides may also be used to identify appetite suppressants, to
CC trap odours of a specific type, as biosensors for the detection of
CC explosives, drugs, perfumes or pollutants, and in apiculture to modify
CC the behaviour of bees, for example, to increase the production of royal
CC jelly.
XX
SQ Sequence 1359 BP; 312 A; 329 C; 333 G; 385 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1359;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaat 23
||||||||||||||
Db 162 CAAGTCACAGAAT 150

RESULT 45
AAZ59716/c
ID AAZ59716 standard; DNA; 1404 BP.
XX
AC AAZ59716;
XX
DT 19-APR-2000 (first entry)
XX
DE DNA encoding a mutant phytase-10, phytase-10-thermo[3]-Q50T-K91A.
XX
KW Phytase; myo-inositol hexakisphosphate phosphohydrolase; stabilisation;
KW thermostable; animal feed; monogastric animal; phytate phosphorus;
KW phosphate availibility; consensus; mutant; ds.
XX
OS Aspergillus terreus 9A1.
OS Aspergillus terreus cbs16.46.
OS Aspergillus niger var. awamori.
OS Aspergillus niger str. NRRL3135.
OS Aspergillus fumigatus ATCC13073.
OS Aspergillus fumigatus ATCC32722.
OS Aspergillus fumigatus ATCC58128.
OS Aspergillus fumigatus ATCC26906.
OS Aspergillus fumigatus ATCC32239.
OS Emericella nidulans.
OS Talaromyces thermophilus ATCC20186.
OS Myceliophthora thermophila.
OS Paxillus involutus NN005693.
OS Trametes pubescens NN9343.
OS Agrocybe pediades NN009289.
OS Peniophora lycii NN006113.
OS Thermomyces lanuginosa.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1404
FT /*tag= a
FT /product= "Phytase-10-thermo[3-Q50T-K91A"
XX
PN EP969089-A1.
XX
PD 05-JAN-2000.
XX
PF 23-JUN-1999; 99EP-0111949.
XX
PR 29-JUN-1998; 98EP-0111960.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
PI Brugger R, Lehmann M, Wyss M;
XX
DR WPI; 2000-099429/09.
XX
PT New stabilized enzyme formulation, useful for feed compositions for
PT monogastric animals -
XX
PS Example 5; Fig 19; 101pp; English.
XX
CC The invention relates to a novel stabilised dry or liquid enzyme
CC formulation, comprising phytase (myo-inositol hexakisphosphate
CC phosphohydrolase) and one or more stabilising agents including
CC xylitol or ribitol; polyethylene glycols with a molecular weight of 600
CC to 4000 Da, preferably 1000 to 3350 Da; the disodium salts of malonic,
CC glutaric and succinic acid; carboxymethylcellulose; and sodium alginate.
CC The stabilised phytase formulation is used in a method for preparing a
CC feed composition for monogastric animals (e.g., pigs, poultry) and
CC provides a monogastric animal with its dietary requirements of
CC phosphorus. Although a large amount of phosphate is present in animal
CC feed in the form of phytate phosphorus, monogastric animals are unable
CC to utilise this form of phosphate, resulting in the addition of extra
CC phosphate to the feed of such animals. Phytase enhances the nutritional
CC value of plant material without the need for adding additional phosphate
CC to the feed. The level of phosphate pollution in the environment is
CC reduced by adding phytase to animal feed, as the animal can make use of
CC the inorganic phosphate liberated from phytate phosphorus using the
CC enzyme. The phytase formulation of the invention has an improved
CC thermostability and can therefore remain stable during long-term storage
CC and can withstand feed processing methods such as extrusion, expansion
CC and pelleting. The present sequence represents DNA encoding a mutant
CC phytase-10 consensus sequence, phytase-10-thermo[3]-Q50T-K91A, which has
CC a temperature optimum and melting point 4 degrees Celsius higher than
CC that of phytase-10 (AAY69566). Its specific activity with phytate as a
CC substrate is also strongly increased.
XX
SQ Sequence 1404 BP; 329 A; 309 C; 308 G; 458 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1404;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaat 23
||||||||||||||
Db 852 CAAGTCACAGAAT 840

Search completed: October 9, 2001, 15:51:43
Job time: 15064 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:15:24 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 7928

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c 5	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli
6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appl
c 7	14	56.0	1050	6	5498529-7	Patent No. 5498529
c 8	14	56.0	1677	2	US-08-372-652-14	Sequence 14, Appl
c 9	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appl
c 10	14	56.0	1787	2	US-08-372-652-6	Sequence 6, Appli
c 11	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appli
c 12	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appli
c 13	14	56.0	2585	4	US-09-008-697A-7	Sequence 7, Appli
c 14	14	56.0	3083	1	US-07-718-575-5	Sequence 5, Appli
c 15	14	56.0	3083	1	US-08-481-206-5	Sequence 5, Appli
c 16	14	56.0	3083	2	US-08-486-269A-5	Sequence 5, Appli
c 17	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appli
c 18	13	52.0	1662	3	US-09-066-544-1	Sequence 1, Appli
c 19	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appli
20	13	52.0	2091	1	US-08-765-081-4	Sequence 4, Appli
21	13	52.0	2091	3	US-09-098-082-4	Sequence 4, Appli
22	13	52.0	2091	5	PCT-US95-06994-4	Sequence 4, Appli
c 23	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appli
c 24	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appli
c 25	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appli
c 26	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appli
c 27	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appli
c 28	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appli
c 29	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appli
c 30	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appli
c 31	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appli
c 32	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appli
c 33	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appli
c 34	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appli
c 35	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
c 36	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
37	13	52.0	8041	1	US-08-765-081-1	Sequence 1, Appli
38	13	52.0	8041	3	US-09-098-082-1	Sequence 1, Appli
39	13	52.0	8041	5	PCT-US95-06994-1	Sequence 1, Appli
c 40	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appli
c 41	12	48.0	24	3	US-09-014-065-15	Sequence 15, Appl
42	12	48.0	255	5	PCT-US93-03077-7	Sequence 7, Appli
c 43	12	48.0	279	1	US-08-047-041A-17	Sequence 17, Appl
c 44	12	48.0	308	4	US-09-232-522-3	Sequence 3, Appli
c 45	12	48.0	311	1	US-08-047-041A-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7
```

```
  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
```

```
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792
;  GENERAL INFORMATION:
;    APPLICANT:  Ifuku, Ohji
;    APPLICANT:  Haze, Shinitiro
;    APPLICANT:  Kishimoto, Jiro
;    APPLICANT:  Nakahama, Kazuo
;    TITLE OF INVENTION:  BIOTIN OPERON
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wegner, Cantor, Mueller & Player
;      STREET:  1233 20th Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-8218
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/944,259
;      FILING DATE:  19920914
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cantor, Herbert I.
;      REGISTRATION NUMBER:  24,392
;      REFERENCE/DOCKET NUMBER:  P-450-23557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-0400
;      TELEFAX:  202-835-0605
;      TELEX:  440706
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
US-07-944-259-2


  Query Match             68.0%;  Score 17;  DB 2;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.23;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gtcgcaagtcacagaat 23
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1


RESULT   6
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0323
;  CURRENT APPLICATION NUMBER: US/09/166,203A
;  CURRENT FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 60
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48


  Query Match             60.0%;  Score 15;  DB 2;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   7
5498529-7/c
;Patent No. 5498529
;    APPLICANT: BERKA, THOMAS R.;FORNWALD, JAMES A.;GORNIAK,
;JOSELINA G.;ROSENBERG, MARTIN;STRICKLER, JAMES E.;TAYLOR, DEAN P.
;    TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
;STREPTOMYCES
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,506
;      FILING DATE: 03-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 894,167
;      FILING DATE: 03-JUN-1992
;      APPLICATION NUMBER: 346,119
;      FILING DATE: 16-FEB-1989
;      APPLICATION NUMBER: 897,245
;      FILING DATE: 18-AUG-1986
;SEQ ID NO:7:
;      LENGTH: 1050
5498529-7


  Query Match             56.0%;  Score 14;  DB 6;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db     251 AAGTCACAGAATTA 238


RESULT   8
US-08-372-652-14/c
; Sequence 14, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
```

```
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1677 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-372-652-14


  Query Match             56.0%;  Score 14;  DB 2;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT   9
PCT-US95-16311-14/c
; Sequence 14, Application PC/TUS9516311
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16311
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1677 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16311-14


  Query Match             56.0%;  Score 14;  DB 5;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  10
US-08-372-652-6/c
; Sequence 6, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1787 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-372-652-6
```

```
  Query Match             56.0%;  Score 14;  DB 2;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  11
PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16311
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1787 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16311-6


  Query Match             56.0%;  Score 14;  DB 5;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  12
US-08-496-631-1/c
; Sequence 1, Application US/08496631
; Patent No. 5728548
;  GENERAL INFORMATION:
;    APPLICANT:  Bowman, Michael
;    TITLE OF INVENTION:  STEROID RECEPTOR RR1
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/496,631
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2233 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  363..1778
US-08-496-631-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 8.4;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db    1055 TTGTCGCAAGTCAC 1042


RESULT  13
US-09-008-697A-7/c
; Sequence 7, Application US/09008697A
; Patent No. 6197504
;  GENERAL INFORMATION:
;    APPLICANT:  Chow, King Lau
;    TITLE OF INVENTION:  USES OF MAB-21
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,697A
;      FILING DATE:  January 19, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chan, Albert Wai-Kit
;      REGISTRATION NUMBER:  36,479
;      REFERENCE/DOCKET NUMBER:  50752-A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2585 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-008-697A-7


  Query Match             56.0%;  Score 14;  DB 4;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1058 CAAGTCACAGAATT 1045


RESULT  14
US-07-718-575-5/c
; Sequence 5, Application US/07718575
; Patent No. 5202257
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/718,575
;      FILING DATE:  19910813
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-07-718-575-5


  Query Match             56.0%;  Score 14;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 cattgtcgcaagtc 16
          ||||||||||||||
Db   3034 CATTGTCGCAAGTC 3021


RESULT  15
US-08-481-206-5/c
; Sequence 5, Application US/08481206
; Patent No. 5739291
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,206
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/013,767
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-08-481-206-5


  Query Match             56.0%;  Score 14;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 cattgtcgcaagtc 16
          ||||||||||||||
Db   3034 CATTGTCGCAAGTC 3021
```

```
RESULT  16
US-08-486-269A-5/c
; Sequence 5, Application US/08486269A
; Patent No. 5945509
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann, Stephen F.
;    APPLICANT:  Boulter, James R.
;    APPLICANT:  Hollmann, Michael
;    APPLICANT:  Bettler, Bernhard
;    APPLICANT:  Jensen, Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,269A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/013,767
;      FILING DATE:  04-FEB-1993
;      APPLICATION NUMBER:  07/718,575
;      FILING DATE:  21-JUN-1991
;      APPLICATION NUMBER:  PCT/US90/06153
;      FILING DATE:  25-OCT-1990
;      APPLICATION NUMBER:  07/428,116
;      FILING DATE:  27-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9986
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  167...2830
;      OTHER INFORMATION:
US-08-486-269A-5


  Query Match             56.0%;  Score 14;  DB 2;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 cattgtcgcaagtc 16
          ||||||||||||||
Db   3034 CATTGTCGCAAGTC 3021


RESULT  17
```

```
US-08-651-572-1/c
; Sequence 1, Application US/08651572
; Patent No. 5789228
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-651-572-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    974 AAGTCACAGAATT 962


RESULT  18
US-09-066-544-1/c
; Sequence 1, Application US/09066544
; Patent No. 6001984
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
```

```
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,544
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-066-544-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  19
US-08-951-086-1/c
; Sequence 1, Application US/08951086
; Patent No. 6074867
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,086
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-951-086-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  20
US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-08-765-081-4


  Query Match             52.0%;  Score 13;  DB 1;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  21
US-09-098-082-4
; Sequence 4, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-09-098-082-4


  Query Match             52.0%;  Score 13;  DB 3;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  22
```

```
PCT-US95-06994-4
; Sequence 4, Application PC/TUS9506994
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
PCT-US95-06994-4


  Query Match             52.0%;  Score 13;  DB 5;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  23
US-08-484-993B-1/c
; Sequence 1, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
```

```
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-993B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  24
US-08-484-158B-1/c
; Sequence 1, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-158B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  25
US-08-484-596A-1/c
; Sequence 1, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-596A-1

  Query Match '        '  52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  26
US-08-480-150A-1/c
; Sequence 1, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  12..2153
US-08-480-150A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  27
US-08-458-731-1/c
; Sequence 1, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-458-731-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  28
US-08-149-223A-1/c
; Sequence 1, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
```

```
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-149-223A-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  29
US-08-484-993B-9/c
; Sequence 9, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-993B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  30
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
```

```
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-158B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  31
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-596A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  32
US-08-480-150A-9/c
; Sequence 9, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-480-150A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  33
US-08-458-731-9/c
; Sequence 9, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-458-731-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  34
US-08-149-223A-9/c
; Sequence 9, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-149-223A-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
```

```
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  35
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
US-08-743-637B-35


  Query Match             52.0%;  Score 13;  DB 2;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    1911 AAGTCACAGAATT 1899


RESULT  36
US-08-526-840B-35/c
; Sequence 35, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
US-08-526-840B-35


  Query Match             52.0%;  Score 13;  DB 3;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    1911 AAGTCACAGAATT 1899


RESULT  37
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
```

```
;       COMPUTER:  IBM PC compatible/Pentium
;       OPERATING SYSTEM:  MS-Window 3.1
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/765,081
;       FILING DATE:  March 26, 1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Shelton, Dennis K.
;       REGISTRATION NUMBER:  26,997
;       REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
US-08-765-081-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  38
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;     APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;     TITLE OF INVENTION:  Escherichia Coli O157:H7 Epithelial Adhesin
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;       STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;       COMPUTER:  IBM PC compatible/Pentium II
;       OPERATING SYSTEM:  MS-Windows 95
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/098,082
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/765,081
;       FILING DATE:  March 26, 1997
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sheiness, Diana K.
;       REGISTRATION NUMBER:  35,356
;       REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
US-09-098-082-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  39
PCT-US95-06994-1
; Sequence 1, Application PC/TUS9506994
;  GENERAL INFORMATION:
;     APPLICANT:  Children's Hospital & Medical Center
;     APPLICANT:  University of Washington
;     APPLICANT:  Washington State University Research Foundation
;     APPLICANT:  TARR, PHILLIP I
;     APPLICANT:  BILGE, SIMA S
;     APPLICANT:  BESSER, THOMAS E
;     APPLICANT:  VARY JR, JAMES C
;     TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;       STREET:  SUITE 2800, 1420 FIFTH AVENUE
;       CITY:  SEATTLE
;       STATE:  WASHINGTON
;       COUNTRY:  USA
;       ZIP:  WA 98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  07-JUN-95
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  24-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BRODERICK, THOMAS F
;       REGISTRATION NUMBER:  31,332
;       REFERENCE/DOCKET NUMBER:  CHOR-18591
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 682 8100
;       TELEFAX:  (206) 224 0779
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
PCT-US95-06994-1


  Query Match             52.0%;  Score 13;  DB 5;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  40
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  72928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db   26365 AAGTCACAGAATT 26353


RESULT  41
US-09-014-065-15/c
; Sequence 15, Application US/09014065
; Patent No. 6033854
; GENERAL INFORMATION:
;  APPLICANT: Kurnit, David M.
;  APPLICANT: Chiang, Pei-Wen
;  APPLICANT: Wang, Chang-Ning J.
;  TITLE OF INVENTION: METHOD FOR DETERMINING THE COPY NUMBER OF A NUCLEIC ACID SEQUE
;  FILE REFERENCE: 06498/004001
;  CURRENT APPLICATION NUMBER: US/09/014,065
;  CURRENT FILING DATE: 1998-01-27
;  EARLIER APPLICATION NUMBER: US 08/434,474
;  EARLIER FILING DATE: 1995-05-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 24
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-014-065-15


  Query Match             48.0%;  Score 12;  DB 3;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacaga 21
           ||||||||||||
Db      12 GCAAGTCACAGA 1


RESULT  42
PCT-US93-03077-7
; Sequence 7, Application PC/TUS9303077
;  GENERAL INFORMATION:
;    APPLICANT:  Board of Regents, The Universityof Texas System
;    APPLICANT:  Gaynor, Richard B.
;    APPLICANT:  Wu, Foon Kin
;    TITLE OF INVENTION:  PROTEIN CELLULAR FACTOR USEFUL FOR
;    TITLE OF INVENTION:  REGULATING GENE EXPRESSION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03077
;      FILING DATE:  19930331
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/862,025
;      FILING DATE:  April 2, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kammerer, Patricia A.
;      REGISTRATION NUMBER:  29,775
;      REFERENCE/DOCKET NUMBER:  UTFD270PCT
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  713-787-1540
;       TELEFAX:  713-749-2679
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  255 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
PCT-US93-03077-7


  Query Match             48.0%;  Score 12;  DB 5;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatt 24
           ||||||||||||
Db     206 AGTCACAGAATT 217


RESULT  43
US-08-047-041A-17/c
; Sequence 17, Application US/08047041A
; Patent No. 5527676
;  GENERAL INFORMATION:
;     APPLICANT:  Vogelstein, Bert
;     APPLICANT:  Baker, Suzanne J.
;     APPLICANT:  Fearon, Eric R.
;     APPLICANT:  Nigro, Janice M.
;     TITLE OF INVENTION:  Detection of Loss of the Wild-Type p53
;     TITLE OF INVENTION:  Gene
;     NUMBER OF SEQUENCES:  28
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Banner & Allegretti, Ltd.
;       STREET:  1001 G Street, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20001.4597
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/047,041A
;       FILING DATE:  22-MAR-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/928,661
;       FILING DATE:  17-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/446,584
;       FILING DATE:  06-DEC-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/330,566
;       FILING DATE:  29-MAR-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kagan, Sarah A.
;       REGISTRATION NUMBER:  32,141
;       REFERENCE/DOCKET NUMBER:  01107.42917
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  202-508-9100
;       TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  279 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:  exon 4
;     PUBLICATION INFORMATION:
;       AUTHORS:  Lamb,
;       JOURNAL:  Mol. Cell. Biol.
;       VOLUME:  6
;       ISSUE:  5
;       PAGES:  1379-1385
;       DATE:  1986
US-08-047-041A-17


  Query Match             48.0%;  Score 12;  DB 1;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacaga 21
           ||||||||||||
Db     276 GCAAGTCACAGA 265


RESULT  44
US-09-232-522-3/c
; Sequence 3, Application US/09232522
; Patent No. 6225117
;  GENERAL INFORMATION:
;     APPLICANT:  Gately, Maurice K.
;     APPLICANT:  Presky, David H.
;     TITLE OF INVENTION:  ANTIBODIES AGAINST HUMAN IL-12
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Hoffmann-La Roche Inc.
;       STREET:  340 Kingsland Street
;       CITY:  Nutley
;       STATE:  New Jersey
;       COUNTRY:  United States
;       ZIP:  07110-1199
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/232,522
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Buchholz, Briana C.
;       REGISTRATION NUMBER:  39,123
;       REFERENCE/DOCKET NUMBER:  CD 1048P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  973-235-6208
;       TELEFAX:  973-235-2363
;  INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  308 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  mouse
;       CELL TYPE:  Hybridoma
;       CELL LINE:  HIL-12F3-20E11
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..306
US-09-232-522-3
```

```
  Query Match             48.0%;  Score 12;  DB 4;  Length 308;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatt 24
           ||||||||||||
Db     246 AGTCACAGAATT 235


RESULT  45
US-08-047-041A-4/c
; Sequence 4, Application US/08047041A
; Patent No. 5527676
;  GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Baker, Suzanne J.
;    APPLICANT:  Fearon, Eric R.
;    APPLICANT:  Nigro, Janice M.
;    TITLE OF INVENTION:  Detection of Loss of the Wild-Type p53
;    TITLE OF INVENTION:  Gene
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001.4597
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/047,041A
;      FILING DATE:  22-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/928,661
;      FILING DATE:  17-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/446,584
;      FILING DATE:  06-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/330,566
;      FILING DATE:  29-MAR-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.42917
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  311 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  exon 4
;    PUBLICATION INFORMATION:
;      AUTHORS:  Buchman, V. L.
;      TITLE:  A variation in the sturcture of the
;      TITLE:  protein-coding region of the human p53 gene
;      JOURNAL:  Gene
;      VOLUME:  70
;      PAGES:  245-252
;      DATE:  1988
US-08-047-041A-4


  Query Match             48.0%;  Score 12;  DB 1;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacaga 21
           ||||||||||||
Db     293 GCAAGTCACAGA 282


Search completed: October  9, 2001, 15:55:17
Job time: 13193 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:27:58 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 467815

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	68.0	357	140	BE800706	BE800706 sq97c06.y
2	17	68.0	388	120	AW781150	AW781150 s189g06.y
c 3	17	68.0	751	138	BE658717	BE658717 GM700007A
4	16	64.0	299	134	BB450863	BB450863 BB450863
c 5	16	64.0	418	150	BF556014	BF556014 UI-R-A1-d
c 6	15	60.0	214	121	AW871559	AW871559 ra79d04.y
7	15	60.0	233	139	BE763893	BE763893 RC4-NT005
8	15	60.0	253	139	BE763892	BE763892 RC4-NT005
9	15	60.0	256	5	AA317110	AA317110 EST19001
c 10	15	60.0	261	191	Z45844	Z45844 HSCZVE041 n
c 11	15	60.0	263	126	BB141316	BB141316 BB141316
12	15	60.0	263	169	BF745516	BF745516 CM2-BT082
c 13	15	60.0	282	169	BF800965	BF800965 QV4-CI010
14	15	60.0	286	6	AA344205	AA344205 EST50086
c 15	15	60.0	292	256	B53227	B53227 CIT-HSP-200
c 16	15	60.0	317	169	BF743951	BF743951 CM2-BT082
c 17	15	60.0	327	188	T30901	T30901 EST24536 Hu
c 18	15	60.0	335	169	BF743964	BF743964 CM2-BT082
c 19	15	60.0	346	169	BF743954	BF743954 CM2-BT082
c 20	15	60.0	349	154	BG466199	BG466199 EtESTed25
21	15	60.0	353	10	AA648449	AA648449 ns22f10.s
22	15	60.0	370	188	T07051	T07051 EST04940 Fe
23	15	60.0	402	23	AI694429	AI694429 wd83f06.x
24	15	60.0	404	17	AI203847	AI203847 qf75h03.x
25	15	60.0	405	168	BF726110	BF726110 by01c01.x
c 26	15	60.0	406	153	BG413230	BG413230 EtESTed18
27	15	60.0	416	104	AI942464	AI942464 wp80c05.x
28	15	60.0	420	3	AA172188	AA172188 zp29f06.s
c 29	15	60.0	427	224	AQ142435	AQ142435 HS_3064_A
c 30	15	60.0	429	121	AW870882	AW870882 ra51h09.y
31	15	60.0	454	224	AQ086771	AQ086771 HS_2175_B
32	15	60.0	455	14	AA973493	AA973493 oo45e05.s
c 33	15	60.0	456	24	AI756736	AI756736 EtESTea24
34	15	60.0	457	19	AI361386	AI361386 qy47f01.x
c 35	15	60.0	464	155	BG562271	BG562271 EtESTee05
36	15	60.0	471	22	AI635259	AI635259 tz79h07.x
37	15	60.0	485	172	BG020076	BG020076 dc47c07.x
c 38	15	60.0	487	154	BG516037	BG516037 EtESTed48
c 39	15	60.0	489	169	BF775935	BF775935 286217 MA
40	15	60.0	498	243	AZ398330	AZ398330 1M0163N15
41	15	60.0	502	12	AA846453	AA846453 aj56b03.s
42	15	60.0	502	23	AI672463	AI672463 wa03e05.x
43	15	60.0	503	22	AI634722	AI634722 tz09e04.x
44	15	60.0	512	115	AW418528	AW418528 xr32c07.x
45	15	60.0	512	233	AQ801291	AQ801291 HS_2088_A

ALIGNMENTS

RESULT 1
BE800706
LOCUS BE800706 357 bp mRNA EST 20-SEP-2000
DEFINITION sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1091 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION BE800706
VERSION BE800706.1 GI:10231818
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 357)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.
FEATURES Location/Qualifiers
source 1. .357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The Clark NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). The library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."
BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN

Query Match 68.0%; Score 17; DB 140; Length 357;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       9 cgcaagtcacagaatta 25
           |||||||||||||||||
Db     138 CGCAAGTCACAGAATTA 154
```

RESULT 2
AW781150
LOCUS AW781150 388 bp mRNA EST 12-MAY-2000
DEFINITION s189g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1037-1235 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION AW781150
VERSION AW781150.1 GI:7795753
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 388)

```
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1037-1235"
                     /clone_lib="Gm-c1037"
                     /tissue_type="fully expanded leaves of greenhouse grown
                     plants"
                     /dev_stage="2 week old"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from fully
                     expanded leaves of greenhouse grown plants that were 2
                     weeks old. The library was prepared using the Life
                     Technologies pSuperScript cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a
                     poly(dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the  blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli Electro- Max DH10B host
                     cells. This library was constructed in the laboratory of
                     Dr. Lila Vodkin by Anu Khanna at the University of
                     Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"
BASE COUNT      123 a     67 c     92 g    106 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 120;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 5.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 cgcaagtcacagaatta 25
           |||||||||||||||||
Db     334 CGCAAGTCACAGAATTA 350


RESULT   3
BE658717/c
LOCUS       BE658717      751 bp    mRNA            EST       06-SEP-2000
DEFINITION  GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',
            mRNA sequence.
ACCESSION   BE658717
VERSION     BE658717.1  GI:9984609
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW156684 corresponding to Gm-c1015-2632 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-2561"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID of
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
BASE COUNT      193 a    171 c    131 g    226 t     30 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 138;  Length 751;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 cgcaagtcacagaatta 25
           |||||||||||||||||
Db     112 CGCAAGTCACAGAATTA 96


RESULT   4
BB450863
LOCUS       BB450863      299 bp    mRNA            EST       21-JUL-2000
DEFINITION  BB450863 RIKEN full-length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130006I17 3', mRNA sequence.
ACCESSION   BB450863
VERSION     BB450863.1  GI:9311898
KEYWORDS    EST.
SOURCE      house mouse.
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D130006I17"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal ganglion"
                     /tissue_type="spinal ganglion"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       75 a      72 c      45 g     107 t
ORIGIN

  Query Match           64.0%;  Score 16;  DB 134;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 cgcaagtcacagaatt 24
           ||||||||||||||||
Db       3 CGCAAGTCACAGAATT 18


RESULT   5
BF556014/c
LOCUS       BF556014       418 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-A1-dz-d-02-0-UI.r1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-dz-d-02-0-UI 5', mRNA sequence.
ACCESSION   BF556014
VERSION     BF556014.1  GI:11665744
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1771157
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A1-dz-d-02-0-UI"
                     /clone_lib="UI-R-A1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                     library is a subtracted library derived from the UI-R-A0
                     library. The UI-R-A0 library consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     rat placenta, adult lung, brain, liver, kidney, heart,
                     spleen, ovary, and muscle. The tag is a string of 3-5
                     nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-A1) was constructed as follows: PCR
                     amplified cDNA inserts from a pool of approximately 3,840
                     UI-R-A0 clones from which 3' ESTs had been derived was
                     used as a driver in a hybridization with the UI-R-A0
                     library in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library) was
                     purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     UI-R-A1 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research 6:
                     791-806, 1996)"
BASE COUNT      121 a      97 c     117 g      83 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 150;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtc 16
           ||||||||||||||||
Db     110 GACATTGTCGCAAGTC 95


RESULT   6
AW871559/c
LOCUS       AW871559      214 bp    mRNA            EST       05-APR-2001
DEFINITION  ra79d04.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
            cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION   AW871559
VERSION     AW871559.1  GI:8005612
KEYWORDS    EST.
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 214)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Uma Rao and David Bird
            (david_bird@ncsu.edu) at North Carolina State University. DNA
            Sequencing by: Washington University Genome Sequencing Center St.
            Louis.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .214
                     /organism="Meloidogyne incognita"
                     /db_xref="taxon:6306"
                     /clone_lib="Bird-Rao Meloidogyne incognita J2"
                     /dev_stage="enriched for 2nd stage juveniles"
                     /lab_host="XLORL"
                     /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; Oligo (dT) primed library.  cDNA was
                     contructed and cloned  unidirectionally into the vector
                     within the 5' EcoRI and  3' XhoI sites.  This library was
                     constructed by Dr. Uma Rao  and Dr. David Bird at North
                     Carolina state University."
BASE COUNT       59 a     52 c     19 g     84 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 121;  Length 214;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     184 CAAGTCACAGAATTA 170
```

```
RESULT   7
BE763893
LOCUS       BE763893      233 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763893
VERSION     BE763893.1  GI:10193817
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-NT0054-120
            600-015-g05&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 233.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: pucl8; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a     27 c     52 g     66 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 139;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db      94 CAAGTCACAGAATTA 108


RESULT   8
BE763892
LOCUS       BE763892      253 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763892
VERSION     BE763892.1  GI:10193816
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC4-NT0054-120
            600-015-g03&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       95 a     31 c     56 g     71 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 139;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     114 CAAGTCACAGAATTA 128
```

```
RESULT   9
AA317110
LOCUS       AA317110       256 bp    mRNA            EST       19-APR-1997
DEFINITION  EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA317110
VERSION     AA317110.1  GI:1969448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):117571"
                     /db_xref="taxon:9606"
                     /clone_lib="Lung"
                     /dev_stage="adult"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI"
BASE COUNT       73 a     57 c     48 g     75 t      3 others
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 5;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     189 CAAGTCACAGAATTA 203
```

```
RESULT  10
Z45844/c
LOCUS       Z45844         261 bp    mRNA            EST       14-NOV-1994
DEFINITION  HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone
            c-zve04, mRNA sequence.
ACCESSION   Z45844
VERSION     Z45844.1  GI:575078
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
```

```
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-zve04
            Seq primer: (-21)M13_universal.
FEATURES              Location/Qualifiers
     source           1. .261
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="c-zve04"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
                      /tissue_type="total brain"
                      /dev_stage="3 months old"
                      /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                      Site_2: NotI; sex=Female; dev_stage=3 months old;
                      isolate=muscular atrophy patient;  tissue_type=total brain
                      ;  total mRNA was oligo-(dT) primed and directionally
                      cloned 5' -> 3' into the  HindIII -> NotI sites of the
                      lafmid BA vector. Clone library from B.Soares,  Psychiatry
                      Dept. Columbia University, USA.  Normalization_method:
                      Bento Soares, P.N.A.S in press"
BASE COUNT       77 a     56 c     31 g     97 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 191;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     138 CAAGTCACAGAATTA 124


RESULT  11
BB141316/c
LOCUS       BB141316      263 bp    mRNA            EST       28-JUN-2000
DEFINITION  BB141316 RIKEN full-length enriched, adult female vagina Mus
            musculus cDNA clone 9930007O04 3', mRNA sequence.
ACCESSION   BB141316
VERSION     BB141316.1  GI:8796253
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .263
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="9930007O04"
                      /clone_lib="RIKEN full-length enriched, adult female
                      vagina"
                      /sex="female"
                      /tissue_type="vagina"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. cDNA went through one round of normalization
                      to Rot = 10.0 and subtraction to Rot = 185.0. Second
                      strand cDNA was prepared with the primer adapter of
                      sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC
                      3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                      modified pBluescript KS(+) after bulk excision from Lambda
                      FLC I."
BASE COUNT       69 a     56 c     43 g     95 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 126;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db      62 CAAGTCACAGAATTA 48


RESULT  12
BF745516
LOCUS       BF745516      263 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-021100-469-h05 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF745516
VERSION     BF745516.1  GI:12072192
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
```

```
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            021100-469-h05&t3=2000-11-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 129
            High quality sequence stop: 263.
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       77 a     55 c     60 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     217 GCAAGTCACAGAATT 231


RESULT  13
BF800965/c
LOCUS       BF800965      282 bp    mRNA            EST       12-JAN-2001
DEFINITION  QV4-CI0100-231000-457-d12 CI0100 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF800965
VERSION     BF800965.1  GI:12129954
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV4&t2=QV4-CI0100-
            231000-457-d12&t3=2000-10-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CI0100"
                     /dev_stage="Adult"
                     /note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     75 c     67 g     72 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 282;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 attgtcgcaagtcac 18
           ||||||||||||||||
Db      51 ATTGTCGCAAGTCAC 37


RESULT  14
AA344205
LOCUS       AA344205      286 bp    mRNA            EST       21-APR-1997
DEFINITION  EST50086 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA344205
VERSION     AA344205.1  GI:1996443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
```

```
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .286
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):145738"
                      /db_xref="taxon:9606"
                      /clone_lib="Gall bladder I"
                      /sex="female"
                      /dev_stage="adult, 25 yrs"
                      /note="Organ: gall bladder; Vector: pBluescript SK-;
                      Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       83 a      62 c      48 g      91 t       2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     197 CAAGTCACAGAATTA 211


RESULT  15
B53227/c
LOCUS       B53227         292 bp    DNA             GSS       20-JUN-1998
DEFINITION  CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
            sequence.
ACCESSION   B53227
VERSION     B53227.1  GI:2607561
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .292
                      /organism="Homo sapiens"
                      /db_xref="GDB:7040812"
                      /db_xref="taxon:9606"
                      /clone="2009H2"
                      /clone_lib="CIT-HSP"
                      /sex="Male"
                      /cell_type="Sperm"
                      /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                      HindIII"
BASE COUNT       58 a      64 c      90 g      80 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 256;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     248 CAAGTCACAGAATTA 234


RESULT  16
BF743951/c
LOCUS       BF743951       317 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743951
VERSION     BF743951.1  GI:12070627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-c02&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 317.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT       96 a      71 c      49 g     101 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      45 GCAAGTCACAGAATT 31
```

```
RESULT  17
T30901/c
LOCUS       T30901         327 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
            sequence.
ACCESSION   T30901
VERSION     T30901.1  GI:612999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
            Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
            Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC10632
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):100882"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Brain"
                     /note="Organ: brain"
BASE COUNT       89 a     75 c     47 g    115 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 188;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     295 CAAGTCACAGAATTA 281


RESULT  18
BF743964/c
LOCUS       BF743964       335 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743964
VERSION     BF743964.1  GI:12070640
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-h04&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 40
            High quality sequence stop: 335.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      101 a     76 c     55 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 335;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      61 GCAAGTCACAGAATT 47


RESULT  19
BF743954/c
LOCUS       BF743954       346 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743954
VERSION     BF743954.1  GI:12070630
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
```

```
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-d07&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 43
            High quality sequence stop: 346.
FEATURES              Location/Qualifiers
     source           1. .346
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT      103 a     78 c     54 g    111 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      73 GCAAGTCACAGAATT 59


RESULT  20
BG466199/c
LOCUS       BG466199      349 bp    mRNA            EST       20-MAR-2001
DEFINITION  EtESTed25c05.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to TR:O88349 O88349 LATENT TGF BETA BINDING
            PROTEIN. ;, mRNA sequence.
ACCESSION   BG466199
VERSION     BG466199.1  GI:13395174
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 318.
FEATURES              Location/Qualifiers
     source           1. .349
                      /organism="Eimeria tenella"
                      /strain="LS18"
                      /db_xref="taxon:5802"
                      /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                      /dev_stage="Sporozoite stage"
                      /lab_host="SOLR"
                      /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                      ; Sporozoites were obtained from in vitro sporulated and
                      excysted oocysts of E.tenella grown in chickens. cDNA was
                      synthesized from poly mRNA using an oligo-dT primer
                      containing a XhoI site. Following second strand synthesis,
                      EcoRI adapters were ligated to the cDNA and products were
                      size-selected on sephacryl S500. The cDNA were ligated to
                      EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                      converted to phagemids by mass excision using ExAssist
                      helper phage and E.coli SOLR cell (Stratagene). Clones
                      were selected by negative hybridization against a pool of
                      overrepresented ESTs (N>=10,from 1682 previous reads).
                      Insert sizes range from 1.2-2.9Kb. The library may contain
                      a small percentage of host or bacterial contaminents."
BASE COUNT      75 a    107 c    103 g     64 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 154;  Length 349;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     328 GACATTGTCGCAAGT 314


RESULT  21
AA648449
LOCUS       AA648449      353 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns22f10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
            similar to contains element LTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AA648449
VERSION     AA648449.1  GI:2574878
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Insert Length: 2027   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 334.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1184395"
                      /clone_lib="NCI_CGAP_GCB1"
                      /tissue_type="germinal center B cell"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 49 c 39 g 134 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 249 CAAGTCACAGAATTA 263

RESULT 22
T07051
LOCUS T07051 370 bp mRNA EST 30-JUN-1993
DEFINITION EST04940 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBEC79, mRNA sequence.
ACCESSION T07051
VERSION T07051.1 GI:318200
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE 3,400 expressed sequence tags identify diversity of transcripts from human brain
JOURNAL Nature Genet. 4, 256-267 (1993)
MEDLINE 93364420
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .370
/organism="Homo sapiens"
/db_xref="ATCC (inhost):83698"
/db_xref="taxon:9606"
/clone="HFBEC79"
/clone_lib="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female; oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average inser size."
BASE COUNT 137 a 58 c 68 g 106 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 188; Length 370;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 113 CAAGTCACAGAATTA 127

RESULT 23
AI694429
LOCUS AI694429 402 bp mRNA EST 02-JUN-1999
DEFINITION wd83f06.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338211 3', mRNA sequence.
ACCESSION AI694429
VERSION AI694429.1 GI:4971769
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2338211"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 160 a 60 c 69 g 113 t
ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 402;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 342 CAAGTCACAGAATTA 356

RESULT 24
AI203847
LOCUS AI203847 404 bp mRNA EST 27-JAN-1999
DEFINITION qf75h03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755893 3', mRNA sequence.
ACCESSION AI203847
VERSION AI203847.1 GI:3756453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

```
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 563   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 394.
FEATURES              Location/Qualifiers
     source           1. .404
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1755893"
                      /clone_lib="Soares_testis_NHT"
                      /sex="male"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from mRNA obtained from Clontech Laboratories
                      , Inc., and primed with  a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                      Double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization to Cot5, and was
                      constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      161 a     61 c     69 g    113 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 17;  Length 404;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     342 CAAGTCACAGAATTA 356


RESULT  25
BF726110
LOCUS       BF726110      405 bp    mRNA            EST       05-JAN-2001
DEFINITION  by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION   BF726110
VERSION     BF726110.1  GI:12042021
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 01  row: c  column: 01
            Seq primer: -21M13 forward primer (ABI).
FEATURES              Location/Qualifiers
     source           1. .405
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="by01c01"
                      /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                      BY"
                      /tissue_type="Lens"
                      /dev_stage="Adult"
                      /lab_host="EMDH10B"
                      /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                      from different adults (both approximately 40 years old)
                      together yielded 20ug of total RNA and 150ng mRNA for cDNA
                      library synthesis. A directionally cloned cDNA library in
                      the pCMVSPORT6 vector was constructed at Life Technologies
                      , essentially following the protocols of the SuperScript
                      Plasmid System full details of which are contained in the
                      manufacturer's Instruction manual
                      (http://www.lifetech.com/). First strand synthesis was
                      carried out using a Not I primer-adapter
                      [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                      end inserts were cloned into the Not I/EcoR V sites in the
                      vector. EST analysis was performed on the unamplified
                      library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT      140 a     58 c     47 g    160 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 168;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     249 CAAGTCACAGAATTA 263


RESULT  26
BG413230/c
LOCUS       BG413230      406 bp    mRNA            EST       13-MAR-2001
DEFINITION  EtESTed18g03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG413230
VERSION     BG413230.1  GI:13318783
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .406
                      /organism="Eimeria tenella"
                      /strain="LS18"
                      /db_xref="taxon:5802"
                      /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                      /dev_stage="Sporozoite stage"
                      /lab_host="SOLR"
```

/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10,from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminents."

BASE COUNT 91 a 126 c 114 g 75 t
ORIGIN

Query Match 60.0%; Score 15; DB 153; Length 406;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagt 15
||||||||||||||||
Db 406 GACATTGTCGCAAGT 392

RESULT 27
AI942464
LOCUS AI942464 416 bp mRNA EST 06-SEP-1999
DEFINITION wp80c05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2468072 3', mRNA sequence.
ACCESSION AI942464
VERSION AI942464.1 GI:5707120
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 416)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 410.
FEATURES Location/Qualifiers
source 1. .416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2468072"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 164 a 63 c 73 g 116 t
ORIGIN

Query Match 60.0%; Score 15; DB 104; Length 416;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||||||||||||
Db 331 CAAGTCACAGAATTA 345

RESULT 28
AA172188
LOCUS AA172188 420 bp mRNA EST 09-MAR-1998
DEFINITION zp29f06.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone IMAGE:610883 3', mRNA sequence.
ACCESSION AA172188
VERSION AA172188.1 GI:1751265
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 420)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 934 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 294.
FEATURES Location/Qualifiers
source 1. .420
/organism="Homo sapiens"
/db_xref="GDB:4626142"
/db_xref="taxon:9606"
/clone="IMAGE:610883"
/clone_lib="Stratagene neuroepithelium (#937231)"
/dev_stage="Ntera-2/RA neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2 cells (Ntera-2/c1.D1) induced with Retinoic Acid for 24 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 164 a 63 c 72 g 118 t 3 others
ORIGIN

Query Match 60.0%; Score 15; DB 3; Length 420;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||||||||||||
Db 338 CAAGTCACAGAATTA 352

```
RESULT  29
AQ142435/c
LOCUS       AQ142435      427 bp    DNA             GSS       24-SEP-1998
DEFINITION  HS_3064_A1_A11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3064 Col=21 Row=A, DNA sequence.
ACCESSION   AQ142435
VERSION     AQ142435.1  GI:3533088
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3064  row: A  column: 21
            Class: BAC ends
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3064 Col=21 Row=A"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      137 a     78 c    103 g    108 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 224;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     198 GCAAGTCACAGAATT 184


RESULT  30
AW870882/c
LOCUS       AW870882      429 bp    mRNA            EST       05-APR-2001
DEFINITION  ra51h09.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
            cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION   AW870882
VERSION     AW870882.1  GI:8004935
KEYWORDS    EST.
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Uma Rao and David Bird
            (david_bird@ncsu.edu) at North Carolina State University. DNA
            Sequencing by: Washington University Genome Sequencing Center St.
            Louis.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Meloidogyne incognita"
                     /db_xref="taxon:6306"
                     /clone_lib="Bird-Rao Meloidogyne incognita J2"
                     /dev_stage="enriched for 2nd stage juveniles"
                     /lab_host="XLORL"
                     /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; Oligo (dT) primed library.  cDNA was
                     contructed and cloned  unidirectionally into the vector
                     within the 5' EcoRI and  3' XhoI sites.  This library was
                     constructed by Dr. Uma Rao  and Dr. David Bird at North
                     Carolina state University."
BASE COUNT      127 a     66 c     38 g    198 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 121;  Length 429;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     385 CAAGTCACAGAATTA 371


RESULT  31
AQ086771
LOCUS       AQ086771      454 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2175_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION   AQ086771
VERSION     AQ086771.1  GI:3455988
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2175  row: B  column: 12
            Class: BAC ends
            High quality sequence stop: 454.
```

```
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2175 Col=12 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       188 a     70 c     76 g    117 t      3 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 224;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     156 CAAGTCACAGAATTA 170


RESULT  32
AA973493
LOCUS       AA973493      455 bp    mRNA            EST       17-JUN-1998
DEFINITION  oo45e05.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1569152 3',
            mRNA sequence.
ACCESSION   AA973493
VERSION     AA973493.1  GI:3148673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 553   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1569152"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       175 a     64 c     84 g    132 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 14;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     340 CAAGTCACAGAATTA 354


RESULT  33
AI756736/c
LOCUS       AI756736      456 bp    mRNA            EST       18-JAN-2000
DEFINITION  EtESTea24h07.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1 PRECURSOR ;, mRNA
            sequence.
ACCESSION   AI756736
VERSION     AI756736.1  GI:5150459
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria S5-2 Sporozoite stage"
                     /dev_stage="Sporozoite"
                     /lab_host="SOLR E. coli"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E. tenella grown in chickens.  cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to  the cDNA and products were
                     size-selected on Sephacryl S500.  cDNAs were digested with
                     EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                     Clones were converted to phagemids by mass excision  using
                     ExAssist helper phage and SOLR cells (Stratagene ).
                     Insert  sizes range from 1.2-2.9 kb."
BASE COUNT       100 a    137 c    132 g     87 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 24;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     435 GACATTGTCGCAAGT 421


RESULT  34
```

```
AI361386
LOCUS       AI361386      457 bp    mRNA            EST       15-FEB-1999
DEFINITION  qy47f01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2015161 3',
            mRNA sequence.
ACCESSION   AI361386
VERSION     AI361386.1  GI:4113007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1228   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2015161"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      153 a     73 c     52 g    179 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 19;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     153 CAAGTCACAGAATTA 167


RESULT  35
BG562271/c
LOCUS       BG562271      464 bp    mRNA            EST       10-APR-2001
DEFINITION  EtESTee05a02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG562271
VERSION     BG562271.1  GI:13591269
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT      101 a    141 c    137 g     85 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 155;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     421 GACATTGTCGCAAGT 407


RESULT  36
AI635259
LOCUS       AI635259      471 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294845 3',
            mRNA sequence.
ACCESSION   AI635259
VERSION     AI635259.1  GI:4686589
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1533   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2294845"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      171 a     69 c     53 g    178 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 22;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     235 CAAGTCACAGAATTA 249


RESULT  37
BG020076
LOCUS       BG020076      485 bp    mRNA            EST       24-JAN-2001
DEFINITION  dc47c07.x1 NICHD XGC Emb3 Xenopus laevis cDNA clone IMAGE:3400308
            3' similar to TR:Q9VCA6 Q9VCA6 CG6313 PROTEIN. ;, mRNA sequence.
ACCESSION   BG020076
VERSION     BG020076.1  GI:12476155
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3400308"
                     /clone_lib="NICHD XGC Emb3"
                     /tissue_type="embryo (stages 24-25)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by Life Technologies. Note: This
                     is a Xenopus Gene Collection (XGC) library."
BASE COUNT      117 a    102 c     91 g    165 t     10 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 172;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     437 GCAAGTCACAGAATT 451


RESULT  38
BG516037/c
LOCUS       BG516037      487 bp    mRNA            EST       30-MAR-2001
DEFINITION  EtESTed48c02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG516037
VERSION     BG516037.1  GI:13489363
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT      105 a    147 c    142 g     93 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 154;  Length 487;
```

```
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     451 GACATTGTCGCAAGT 437


RESULT  39
BF775935/c
LOCUS       BF775935      489 bp    mRNA            EST       12-JAN-2001
DEFINITION  286217 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF775935
VERSION     BF775935.1  GI:12123835
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 88  row: K  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .489
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 3BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from marrow, alveolar
                      macrophage, ovary, fetal semitendonosus muscle, and fetal
                      longissimus muscle."
BASE COUNT      147 a     85 c     95 g    162 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 489;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     153 CAAGTCACAGAATTA 139


RESULT  40
AZ398330
LOCUS       AZ398330      498 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0163N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0163N15 R, DNA sequence.
ACCESSION   AZ398330
VERSION     AZ398330.1  GI:10513498
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0163  row: N  column: 15
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 498.
FEATURES              Location/Qualifiers
     source           1. .498
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0163N15"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      187 a     91 c    117 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 243;  Length 498;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     101 GCAAGTCACAGAATT 115


RESULT  41
AA846453
LOCUS       AA846453      502 bp    mRNA            EST       31-DEC-1998
DEFINITION  aj56b03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394285
            3', mRNA sequence.
ACCESSION   AA846453
VERSION     AA846453.1  GI:2932593
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 925   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1394285"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      196 a     67 c     97 g    142 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 12;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     342 CAAGTCACAGAATTA 356


RESULT  42
AI672463
LOCUS       AI672463      502 bp    mRNA            EST       16-DEC-1999
DEFINITION  wa03e05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297024 3',
            mRNA sequence.
ACCESSION   AI672463
VERSION     AI672463.1  GI:4852194
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 535   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2297024"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      198 a     67 c     96 g    141 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 23;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     339 CAAGTCACAGAATTA 353


RESULT  43
AI634722
LOCUS       AI634722      503 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz09e04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2288094 3'
            similar to contains element XTR repetitive element ;, mRNA
            sequence.
ACCESSION   AI634722
VERSION     AI634722.1  GI:4686052
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3719   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .503
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2288094"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      184 a     77 c     56 g    186 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 22;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     236 CAAGTCACAGAATTA 250


RESULT  44
AW418528
LOCUS       AW418528      512 bp    mRNA            EST       09-FEB-2000
DEFINITION  xr32c07.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2761836 3'
            similar to contains element MER32 repetitive element ;, mRNA
            sequence.
ACCESSION   AW418528
VERSION     AW418528.1  GI:6946410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2761836"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT      186 a     79 c     56 g    191 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 115;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     240 CAAGTCACAGAATTA 254


RESULT  45
AQ801291
LOCUS       AQ801291      512 bp    DNA             GSS       09-AUG-1999
DEFINITION  HS_2088_A1_D05_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2088 Col=9 Row=G, DNA sequence.
ACCESSION   AQ801291
VERSION     AQ801291.1  GI:5718623
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2088  row: G  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2088 Col=9 Row=G"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      144 a    117 c     96 g    144 t     11 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 233;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     248 GCAAGTCACAGAATT 262


Search completed: October  9, 2001, 15:15:21
Job time: 13643 sec
```



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:48:27 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 210607

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 BioB gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

	9	25	100.0	5872	9	A93679
	10	25	100.0	5872	9	AR101809
	11	25	100.0	5872	9	AR101810
	12	25	100.0	11022	1	AE000180
	13	21	84.0	7215	3	SMABIO
	14	21	84.0	13501	1	AE005258
	15	21	84.0	297816	2	AP002553
c	16	19	76.0	213732	1	AE001862
	17	17	68.0	2081	94	HAMHPP
c	18	17	68.0	7318	94	AB026497
c	19	17	68.0	69900	66	AC021767
	20	17	68.0	151340	74	AC069132
	21	17	68.0	155204	73	AC068100
c	22	17	68.0	168360	71	AC041048
c	23	17	68.0	176379	72	AC060833
	24	17	68.0	200729	67	AC022781
c	25	16	64.0	1044	53	CNS078T3
	26	16	64.0	8326	8	AF237379
	27	16	64.0	10536	1	AE004050
	28	16	64.0	12242	1	AE005241
	29	16	64.0	38065	3	MLCB1450
	30	16	64.0	39520	78	AC091120
	31	16	64.0	71230	85	AC004508
c	32	16	64.0	85633	85	AC002088
c	33	16	64.0	109973	60	AC007555
	34	16	64.0	110000	84	HSMX1_4
	35	16	64.0	147129	74	AC073052
c	36	16	64.0	155628	71	AC027810
	37	16	64.0	156119	74	AC073247
	38	16	64.0	162043	71	AC040930
c	39	16	64.0	175459	73	AC068538
c	40	16	64.0	183700	86	AC005412
c	41	16	64.0	185492	69	AC025339
	42	16	64.0	202828	72	AC064835
c	43	16	64.0	245499	64	AC016186
	44	16	64.0	262278	2	AP002552
	45	16	64.0	313064	93	HSMX1B

A93679 Sequence 6
AR101809 Sequence
AR101810 Sequence
AE000180 Escherich
D17468 S. marcesce
AE005258 Escherich
AP002553 Escherich
AE001862 Deinococc
M96650 Mesocricetu
AB026497 Mus muscu
AC021767 Mus muscu
AC069132 Homo sapi
AC068100 Homo sapi
AC041048 Homo sapi
AC060833 Homo sapi
AC022781 Mus muscu
AL434413 T7 end of
AF237379 Xenopus l
AE004050 Xylella f
AE005241 Escherich
AL035159 Mycobacte
AC091120 Leishmani
AC004508 Homo sapi
AC002088 Human BAC
AC007555 Homo sapi
Continuation (5 of
AC073052 Homo sapi
AC027810 Homo sapi
AC073247 Homo sapi
AC040930 Homo sapi
AC068538 Homo sapi
AC005412 Homo sapi
AC025339 Homo sapi
AC064835 Homo sapi
AC016186 Homo sapi
AP002552 Escherich
AL442167 Homo sapi

ALIGNMENTS

RESULT 1
AR029499
LOCUS AR029499 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES Location/Qualifiers
source 1. .1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 111 TCGTCAGGTGCAGGTCAGCACGTTG 135

RESULT 2

AR034916
LOCUS AR034916 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES Location/Qualifiers
source 1. .1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 111 TCGTCAGGTGCAGGTCAGCACGTTG 135

RESULT 3
A11530
LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
DEFINITION BioB gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1084)
AUTHORS .
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES Location/Qualifiers
source 1. .1084
/organism="Escherichia coli"
/db_xref="taxon:562"
gene 24. .1064
/gene="BIOB"
CDS 24. .1064
/gene="BIOB"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
PDTDEYYNAAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||

```
Db    134 TCGTCAGGTGCAGGTCAGCACGTTG 158


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 10;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.0027;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    152 TCGTCAGGTGCAGGTCAGCACGTTG 176


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a    1554 c    1631 g    1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 2;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db   2122 TCGTCAGGTGCAGGTCAGCACGTTG 2146


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
```

```
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR",
BASE COUNT     1318 a   1552 c   1695 g   1307 t
```

```
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
```

YINLAFVDASTLGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 227 TCGTCAGGTGCAGGTCAGCACGTTG 251

RESULT 10
AR101809
LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 227 TCGTCAGGTGCAGGTCAGCACGTTG 251

RESULT 11
AR101810
LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES Location/Qualifiers
source 1. .5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 227 TCGTCAGGTGCAGGTCAGCACGTTG 251

RESULT 12
AE000180
LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS .
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"

```
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
     CDS             complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
                     /protein_id="AAC73859.1"
                     /db_xref="GI:1786989"
                     /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                     LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                     AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                     DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                     LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                     NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                     AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                     DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
     promoter        complement(1507. .1535)
                     /note="factor Sigma70; predicted +1 start at 806574"
     gene            complement(1582. .2058)
                     /gene="ybhB"
                     /note="b0773"
     CDS             complement(1582. .2058)
                     /gene="ybhB"
                     /function="orf; Unknown"
                     /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC73860.1"
                     /db_xref="GI:1786990"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     protein_bind    complement(2108. .2124)
                     /note="central position to predicted promoter:85.5"
                     /bound_moiety="RhaS predicted site"
     gene            complement(2117. .3406)
                     /gene="bioA"
                     /note="b0774"
     CDS             complement(2117. .3406)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diaminopelargonic acid synthetase"
                     /protein_id="AAC73861.1"
                     /db_xref="GI:1786991"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     promoter        complement(2193. .2221)
                     /gene="bioA"
                     /note="factor Sigma70; predicted +1 start at 807260"
     protein_bind    3411. .3450
                     /note="central position to bioB promoter: -20"
                     /bound_moiety="BioB documented site"
     protein_bind    3411. .3450
                     /note="central position to predicted promoter: -20;genetic
                     evidence for the site"
                     /bound_moiety="BioB predicted site"
     protein_bind    complement(3411. .3450)
                     /note="central position to predicted promoter:50"
                     /bound_moiety="BioB predicted site"
     protein_bind    complement(3411. .3450)
                     /note="central position to bioA promoter:50"
                     /bound_moiety="BioB documented site"
     promoter        3413. .3441
                     /note="factor Sigma70; promoter bioB; documented +1 at
                     808525"
     promoter        complement(3447. .3473)
                     /note="factor Sigma70; promoter bioA; documented +1 at
                     808515"
     gene            3493. .4533
                     /gene="bioB"
                     /note="b0775"
     CDS             3493. .4533
                     /gene="bioB"
                     /EC_number="2.8.1.-"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis, sulfur insertion?"
                     /protein_id="AAC73862.1"
                     /db_xref="GI:1786992"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            4530. .5684
                     /gene="bioF"
                     /note="b0776"
     CDS             4530. .5684
                     /gene="bioF"
                     /EC_number="2.3.1.47"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAC73863.1"
                     /db_xref="GI:1786993"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL

  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db    3603 TCGTCAGGTGCAGGTCAGCACGTTG 3627


RESULT  13
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
```

```
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
            Phone: 048-433-2545
            Email: nsakurai@ddbj.nig.ac.jp
            Fax:   048-433-2540.
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a   2225 c   2367 g   1276 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 3;  Length 7215;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacg 22
          ||||||||||||||||||||||
Db   2830 CGTCAGGTGCAGGTCAGCACG 2850


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                SYHGGKQDLISVVLSKI"
     gene       6462. .7160
                /gene="Z0990"
     CDS        6462. .7160
                /gene="Z0990"
                /function="orf; Other or unknown (Phage or Prophage
                Related)"
                /note="No significant matches"
                /codon_start=1
                /transl_table=11
                /product="unknown protein encoded by prophage CP-933K"
                /protein_id="AAG55143.1"
                /db_xref="GI:12513758"
                /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                VSLGFQQVRLHPLL"
     gene       complement(7668. .8144)
                /gene="ybhB"
                /note="Z0992"
     CDS        complement(7668. .8144)
                /gene="ybhB"
                /function="orf; Unknown function"
                /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                MG1655: B0773"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG55144.1"
                /db_xref="GI:12513759"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene       complement(8203. .9492)
                /gene="bioA"
                /note="Z0993"
     CDS        complement(8203. .9492)

  Query Match             84.0%;  Score 21;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcac 21
          |||||||||||||||||||||
Db   9689 TCGTCAGGTGCAGGTCAGCAC 9709


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
gene                 2781. .3686
                     /gene="ECs0755"
CDS                  2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
                     transcriptional regulators e.g. glycine cleavage system
                     transcription activator (gcv operon activator) -
                     Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                     identity 31 in 300 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transcription regulator"
                     /protein_id="BAB34178.1"
                     /db_xref="GI:13360214"
                     /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                     EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                     PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                     LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                     NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                     EKSDSPHLQQFREWLLAKSVVPQEM"
gene                 complement(3720. .4322)
                     /gene="ECs0756"
CDS                  complement(3720. .4322)
                     /gene="ECs0756"
                     /note="probable cob(I)alamin adenosyltransferase,similar
                     to cob(I)alamin adenosyltransferases (corrinoid
                     adenosyltransferases) e.g. [Escherichia coli]
                     gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative cob(I)alamin adenosyltransferase"
                     /protein_id="BAB34179.1"
                     /db_xref="GI:13360215"
                     /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                     KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                     QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                     IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene                 complement(4332. .5984)
                     /gene="ECs0757"
CDS                  complement(4332. .5984)
                     /gene="ECs0757"
                     /note="probable fumarate hydratase,similar to fumarate
                     hydratases e.g. fumarate hydratase class I, aerobic
                     (fumarase) - Escherichia coli
                     gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative fumarate hydratase"
                     /protein_id="BAB34180.1"
                     /db_xref="GI:13360216"
                     /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                     ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                     NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                     TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                     GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                     EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                     EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                     DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                     QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                     MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene                 complement(6092. .7372)
                     /gene="ECs0758"
CDS                  complement(6092. .7372)
                     /gene="ECs0758"
                     /note="probable transport protein,similar to
                     glutamate/aspartate transport proteins (proton glutamate
                     sympport proteins) e.g. [Bacillus stearothermophilus]
                     gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                     aa,also similar to C4-dicarboxylate transpor"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transport protein"
                     /protein_id="BAB34181.1"
                     /db_xref="GI:13360217"
                     /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                     TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                     NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                     FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                     VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                     GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                     LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                     AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene                 complement(7533. .7850)
                     /gene="ECs0759"
CDS                  complement(7533. .7850)
                     /gene="ECs0759"
                     /note="unknown"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34182.1"

  Query Match             84.0%;  Score 21;  DB 2;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcac 21
           |||||||||||||||||||||
Db   92421 TCGTCAGGTGCAGGTCAGCAC 92441


RESULT  16
AE001862/c
LOCUS       AE001862   213732 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 1 of 2 of the complete
            chromosome 2.
ACCESSION   AE001862 AE001825
VERSION     AE001862.1  GI:6460468
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
```

```
          Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
          Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE   Direct Submission
  JOURNAL Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .213732
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="2"
     gene            653. .1435
                     /gene="DRA0001"
     CDS             653. .1435
                     /gene="DRA0001"
                     /note="similar to GB:D26185 SP:P37522 PID:467381
                     PID:580906 GB:AL009126 percent identity: 54.96; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chromosome partitioning ATPase, putative, ParA
                     family"
                     /protein_id="AAF12301.1"
                     /db_xref="GI:6460595"
                     /translation="MVSAVKTLTVFNHAGGAGKTSLTLNVGYELARGGLRVLLLDLDP
                     QANLTGWLGISGVTREMTVYPVAVDGQPLPSPVKAFGLDVIPAHVSLAVAEGQMMGRV
                     GAQGRLRRALAEVSGDYDVALIDSPPSLGQLAILAALAADQMIVPVPTRQKGLDALPG
                     LQGALTEYREVRPDLTVALYVPTFYDARRRHDQEVLADLKAHLSPLARPVPQREAVWL
                     DSTAQGAPVSEYAPGTPVHADVQRLTADIAAAIGVAYPGENA"
     gene            1432. .2313
                     /gene="DRA0002"
     CDS             1432. .2313
                     /gene="DRA0002"
                     /note="similar to GB:L34077 GB:X85964 PID:407375
                     PID:757761 PID:1163135 percent identity: 57.34; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chromosome partitioning protein, ParB family"
                     /protein_id="AAF12300.1"
                     /db_xref="GI:6460594"
                     /translation="MTRRRPERRRDLLGLLGETPVDLSQANDIRALPVNELKVGSTQP
                     RRSFDLERLSELAESIRAHGVLQPLLVRSVDGQYEIVAGERRWRAAQLAGLAEVPVVV
                     RQLSNEQARAAALIENLQRDNLNVIDEVDGKLELIALTLGLEREEARKRLMQLLRAVP
                     GDEHEQLDQVFRSMGETWRTFAKNKLRILNWPQPVLEALRAGLPLTLGSVVASAPPER
                     QAELLKLAQNGASRSQLLQALQTPSQTSAVTPEHFAKVLSSKRFLSGLDTPTREALDR
                     WLARMPERVRQAIDEQS"
     gene            complement(2418. .3881)
                     /gene="DRA0003"
     CDS             complement(2418. .3881)
                     /gene="DRA0003"
                     /note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096
                     PID:1789015 percent identity: 69.60; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="succinate-semialdehyde dehydrogenase"
                     /protein_id="AAF12294.1"
                     /db_xref="GI:6460588"
                     /translation="MTTTDLRTTYSSVTRSQAYFDGEWRNAPRNFEVRHPGNGEVIGE
                     VADCTPTDARQAIDAAEVALREWRQVNPYERGKILRRWHDLMFEHKEELAQLMTLEMG
                     KPISETRGEVHYAASFIEWCAEEAGRIAGERINLRFPHKRGLTISEPVGIVYAVTPWN
                     FPAGMITRKAAPALAAGCVMILKPAELSPMTALYLTELWLKAGGPANTFQVLPTNDAS
                     ALTQPFMNDSRVRKLTFTGSTEVGRLLYQQAAGTIKRVSLELGGHAPFLVFDDADLER
                     AASEVVASKFRNSGQTCVCTNRVYVQRGVAEEFIRLLTEKTAALQLGDPFDEATQVGP
                     VVEQAGLDKVQRQVQDALTKGAQATTGGQVSSGLFFQPTVLVDVAPDSLILREETFGP
                     VAPVTIFDTEEEGLRLANDSEYGLAAYAYTRDLGRAFRIAEGLEYGIVGINDGLPSSA
                     APHVPFGGMKNSGVGREGGHWGLEEYLETKFVSLGLS"
     gene            4024. .4668
                     /gene="DRA0004"
     CDS             4024. .4668
                     /gene="DRA0004"
                     /note="similar to GP:2924378 percent identity: 55.29;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12299.1"
                     /db_xref="GI:6460593"
                     /translation="MNSLSSLAAVLLMAGRSTRMGQPKALLPLGGMPLCRHAAQTLVE
                     AGYGEVWAVLPPGETGDAVRAALQGLPLRWVTNPQPAQGLLSSFQAAGEALWAESAAP
                     AVVFTLADMPLVRRDTHQALGTVFGAEQPAAIITRYGEVAAPPTLLRRDLLPRLHEFP
                     PSDHGPRTLLRELGRELKKIDRPAAELLDIDTPEAWAQAQQLFAEVTPGAQDGP"
     gene            complement(4719. .5897)
                     /gene="DRA0005"
     CDS             complement(4719. .5897)
                     /gene="DRA0005"
                     /note="similar to GB:L33464 SP:P47734 PID:496118 percent
                     identity: 74.23; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="alcohol dehydrogenase, zinc-containing"
                     /protein_id="AAF12293.1"
                     /db_xref="GI:6460587"
                     /translation="MKAIVWQGVNKVGVETVPDPTLLLPTDAIVKITSTAICGSDLHL
                     LDGYIPSMETGDILGHEFMGEVVEVGSEVRKLKVGDRVVVPFNIACGVCDPCKRGFFS
                     ACDNSNPNHRMAEALYGATSGGALFGYSHMYGGYAGGQAQYVRVPFADVGPFKIESGL
                     KDEQVLFLTDIFPTGYQAAEQCNIVPGRDVVAVFGAGPVGQFAARSAQMLGAAHVIVI
                     DRVPERLAMAEAAGCQTINYEKEDVLLSLREATGGRGPDHVIDAVGMEAHGHGPGSTM
                     DKVQQNLKLTFDRITALRWAILSCAKGGTVSMPGVYGGLVDKMPIGAAFAKGLIFRMG
                     QTHTHRHIPSLLGRIEAGQIDPSFVITHRASLDQAPELYKTFRDKQDCCIKVVLNPWA
                     "
     gene            complement(5894. .6700)
                     /gene="DRA0006"
     CDS             complement(5894. .6700)
                     /gene="DRA0006"
                     /note="similar to PID:1653488 percent identity: 48.46;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12292.1"
                     /db_xref="GI:6460586"
                     /translation="MDRRSSPQLRTMTNWNREQPAPHAGATGPNSSPGATMQPLERSL
                     TGAVGAALAAAGLRSRSGLEKVVLTGLGAGLIAVAARGQNPLATALKIEQNDEGETVV
                     RDAVTIGKPAEQLYAVWRDLPGLPLLMTHLRSVEVLDDKRSRWTVEAPAPLGTVSWEA
                     ELTADEPGKRIAWRSLPGARIENSGEVLFRPAPGARGTEVVVRLTYRPPGGSAGAVIA
                     RMFNQEPSQQLRDDLMRFKREQELGHAPTAQGQSSGRAALTPPNKRGDLV"
     gene            6839. .7744
                     /gene="DRA0007"
     CDS             6839. .7744
                     /gene="DRA0007"
                     /note="similar to GB:AL009126 percent identity: 62.98;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, ATP-binding protein, putative"
                     /protein_id="AAF12291.1"
                     /db_xref="GI:6460585"
                     /translation="MITTFEQVSKTYGHVTALSDFNLTLRTGELTALLGPNGAGKSTA
                     IGLLLGLSAPSAGQVRVLGADPRRNDVRARIGAMPQESALPAGLTVREAVTLFASFYP
                     APLGVDEALALADLGPVAGRRAAQLSGGQKRRLAFALAVVGDPELLLIDEPTTGMDAQ
                     SRAAFWEAVTGLRARGRTILLTTHYLEEAERTADRVVVMNGGRILADDTPQGLRSGVG
                     GARVSFVSDLVQAELERLPGVSAVQVDAAGRADLRTSVPEALLAALIGSGTTFSDLEV
                     RRATLEEAYLQLTGPQDMTAVTRSA"
     gene            7861. .8652
                     /gene="DRA0008"
     CDS             7861. .8652
                     /gene="DRA0008"
                     /note="similar to GB:AL009126 percent identity: 52.46;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12290.1"
```

```
                     /db_xref="GI:6460584"
                     /translation="MTTLPNVPARRTSALPVLAQLTLGELRRLLRSPMFTVGAIGFPV
                     MFFALFGLPAVQEYGATDPHVGPVILTQFAAYSLLSLALFSFGAAVATERSGGWLRLL
                     RSSPLPVPLYFVSKTLAALAFGAVSLALLYAFAHFAGGVTLPLGLALLLAGKLLLGMI
                     PLVALGLCIGFLASPQAAQILANILSVVMSFASGLFVPLDQLPGFVQQIAPLLPAYHV
                     SQIATNTVSGQTASEPAHWLALAAFTLVFGTLAVWGLKRDESREG"
     gene            8809. .9819
                     /gene="DRA0009"
     CDS             8809. .9819
                     /gene="DRA0009"
                     /note="similar to GB:AL009126 percent identity: 55.80;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12289.1"
                     /db_xref="GI:6460583"
                     /translation="MLFWGTLVAFVAVYVFTFGKRDPDDRWAYVGWLSALLMWALGDW
                     LYGPSALSFLVYGGSLIGWQRRQWVALGGAAINALLMVGRLLTLPERSQGDFIFLLFI
                     LASAYANHATYRGLRARRRLAQVQLEKEKLAADAERERIARDLHDLLGHTLSVIVLKS
                     ELAGKLAEKHPQRAAAEIREVERISREALSEVRAAVQGYRGSGLAAELARAKVALDAA
                     GIRLVITDTLPELPADLESTLAMVLREAVTNVVRHSGASELRLTLRGQANGFQLTLQD
                     DGRGGDAPEGTGLNSMRERLRAMGGTFERDGHTGTRLQAFVPLSAGTAGLKAPELLEK
                     TS"
     gene            9816. .10451
                     /gene="DRA0010"
     CDS             9816. .10451
                     /gene="DRA0010"
                     /note="similar to GB:AL009126 percent identity: 67.84;
```

Query Match 76.0%; Score 19; DB 1; Length 213732;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4 tcaggtgcaggtcagcacg 22
           ||||||||||||||||||||
Db   52231 TCAGGTGCAGGTCAGCACG 52213
```

```
RESULT  17
HAMHPP
LOCUS       HAMHPP       2081 bp    mRNA            ROD       27-APR-1993
DEFINITION  Mesocricetus auratus pregnancy protein (HPP), complete cds.
ACCESSION   M96650
VERSION     M96650.1  GI:191387
KEYWORDS    pregnancy protein; pregnancy protein 60kDa.
SOURCE      Mesocricetus auratus (strain Waterhouse) (library: lambda gt 11)
            female adult, 14 days pregnant liver (maternal) cDNA to mRNA.
  ORGANISM  Mesocricetus auratus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Mesocricetus.
REFERENCE   1  (bases 1 to 2081)
  AUTHORS   Park,C.G.
  TITLE     Cloning and sequencing of a new pregnancy marker protein from
            hamster liver
  JOURNAL   Thesis (1992) Biological Sciences, The Wichita State University
FEATURES             Location/Qualifiers
     source          1. .2081
                     /organism="Mesocricetus auratus"
                     /strain="Waterhouse"
                     /db_xref="taxon:10036"
                     /dev_stage="adult, 14 days pregnant"
                     /sex="female"
                     /tissue_type="liver (maternal)"
                     /tissue_lib="lambda gt 11"
     sig_peptide     68. .130
                     /gene="HPP"
     CDS             68. .1330
                     /gene="HPP"
                     /codon_start=1
                     /product="pregnancy protein 60 kDa"
                     /protein_id="AAA37078.1"
                     /db_xref="GI:191388"
                     /translation="MALRVAFILLVAAFCPAVLCQQDGTLGKETTVQEEQNNETKVDS
                     LTLASINTDFAFSLYKQLALKNPDKNVIFSPFSISFALAFLSLGASSNTLEEILEGLK
                     FNLTETPEADIHRGFGHLLRMLSQPGDQVQVSTSSAMFVEKRLQILAEFKKKARALYQ
                     AEASSTDFQKPHEAKKLINDYVSKKTHGKIKELMSDLDDQTSMVLVNYIYFKGKWKAP
                     FDPDFTYESDFYLDNKRSVKVPMMKHKDLTTPYFRDEELSSYVVEMRYRGNVSAMFIL
                     PDKGRMQQLESSLQPETLRRWKDSLRPRMIHELYVPKFSISTDYSMEGILSQLGIKEV
                     FSSQADLSGITGDKDLRVSKVVHKAVLDVGETGTEAAAATGLRMMGSALIMNPLRLRF
                     NKPFLMIIYNTNTQTPLFMAKVTNPKQN"
     gene            68. .1330
                     /gene="HPP"
     mat_peptide     131. .1327
                     /gene="HPP"
                     /product="pregnancy protein 60 kDa"
BASE COUNT      547 a    572 c    484 g    478 t
ORIGIN
```

Query Match 68.0%; Score 17; DB 94; Length 2081;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5 caggtgcaggtcagcac 21
           |||||||||||||||||
Db     455 CAGGTGCAGGTCAGCAC 471
```

```
RESULT  18
AB026497/c
LOCUS       AB026497     7318 bp    mRNA            ROD       29-APR-2000
DEFINITION  Mus musculus MysPDZ mRNA for myosin containing PDZ domain, complete
            cds.
ACCESSION   AB026497
VERSION     AB026497.1  GI:7416031
KEYWORDS    MysPDZ; myosin containing PDZ domain.
SOURCE      Mus musculus (strain:C57BL6) adult tissue_lib:spleen cDNA to mRNA,
            clone_lib:Lambda ZipLox.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Furusawa,T., Ikawa,S., Yanai,N. and Obinata,M.
  TITLE     Isolation of a novel PDZ-containing myosin from hematopoietic
            supportive bone marrow stromal cell lines
  JOURNAL   Biochem. Biophys. Res. Commun. 270 (1), 67-75 (2000)
  MEDLINE   20200096
REFERENCE   2  (bases 1 to 7318)
  AUTHORS   Furusawa,T., Obinata,M. and Yanai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1999) to the DDBJ/EMBL/GenBank databases. Tadashi
            Furusawa, Institute of Development,Aging and Cancer,Tohoku
            University, Department of Cell Biology; 4-1,Seiryomaci,Aoba-Ku,
            Sendai, Miyagi 980-8575, Japan (E-mail:t-furu@idac.tohoku.ac.jp,
            Tel:81-22-717-8486, Fax:81-22-717-8488)
FEATURES             Location/Qualifiers
     source          1. .7318
                     /organism="Mus musculus"
                     /strain="C57BL6"
                     /db_xref="taxon:10090"
                     /clone_lib="Lambda ZipLox"
                     /dev_stage="adult"
                     /tissue_lib="spleen"
     gene            52. .6159
                     /gene="MysPDZ"
     CDS             52. .6159
                     /gene="MysPDZ"
                     /codon_start=1
                     /product="myosin containing PDZ domain"
                     /protein_id="BAA93660.1"
                     /db_xref="GI:7416032"
                     /translation="MFNLMKKDKDKDGGRKEKKEKKEKKERMSAAELRSLEEMSMRRG
                     FFNLNRSSKRESKTRLEISNPIPIKVASGSDLHLTDIDSDSNRGSIILDSGHLSTASS
                     SDDLKGEEGSFRGSVLQRAAKFGSLAKQNSQMIVKRFSFSQRSRDESASETSTPSEHS
```

```
                     AAPSPQVEVRTLEGQLMQHPGLGIPRPGPRSRVPELVTKRFPADLRLPALVPPPPPAL
                     RELELQRRPTGDFGFSLRRTTMLDRAPEGQAYRRVVHFAEPGAGTKDLALGLVPGDRL
                     VEINGQNVENKSRDEIVEMIRQSGDSVRLKVQPIPELSELSRSWLRTGEGHRREPADA
                     KTEEQIAAEEAWYETEKVWLVHRDGFSLASQLKSEELSLPEGKARVKLDHDGAILDVD
                     EDDIEKANAPSCDRLEDLASLVYLNESSVLHTLRQRYGASLLHTYAGPSLLVLSTRGA
                     PAVYSEKVMHMFKGCRREDMAPHIYAVAQTAYRAMLMSRQDQSIVLLGSSGSGKTTSF
                     QHLVQYLATIAGTSGTKVFSVEKWQALSTLLEAFGNSPTIMNGSATRFSQILSLDFDQ
                     AGQVASASIQTMLLEKLRVARRPASEATFNVFYYLLACGDATLRTELHLNHLAENNVF
                     GIVPLSKPEEKQKAAQQFSKLQAAMKVLAISPEEQKTCWLILASIYHLGAAGATKEAA
                     EAGRKQFARHEWAQKAAYLLGCSLEELSSAIFKHQLKGGTLQRSTSFRQGPEESGLGE
                     GTKLSALECLEGMASGLYSELFTLLISLVNRALKSSQHSLCSMMIVDTPGFQNPEWGG
                     SARGASFEELCHNYAQDRLQRLFHERTFLQELERYKEDNIELAFDDLEPVADDSVAAV
                     DQASHLVRSLAHADEARGLLWLLEEEALVPGATEDALLDRLFSYYGPQEGDKKGQSPL
                     LRSSKPRHFLLGHSHGTNWVEYNVAGWLNYTKQNPATQNAPRLLQDSQKKIISNLFLG
                     RAGSATVLSGSIAGLEGGSQLALRRATSMRKTFTTGMAAVKKKSLCIQIKLQVDALID
                     TIKRSKMHFVHCFLPVAEGWPGEPRSASSRRVSSSSELDLPPGDPCEAGLLQLDVSLL
                     RAQLRGSRLLDAMRMYRQGYPDHMVFSEFRRRFDVLAPHLTKKHGRNYIVVDEKRAVE
                     ELLESLDLEKSSCCLGLSRVFFRAGTLARLEEQRDEQTSRHLTLFQAACRGYLARQHF
                     KKRKIQDLAIRCVQKNIKKNKGVKDWPWWKLFTTVRPLIQVQLSEEQIRNKDEEIQQL
                     RSKLEKVEKERNELRLSSDRLETRISELTSELTDERNTGESASQLLDAETAERLRTEK
                     EMKELQTQYDALKKQMEVMEMEVMEARLIRAAEINGEVDDDDAGGEWRLKYERAVREV
                     DFTKKRLQQELEDKMEVEQQSRRQLERRLGDLQADSDESQRALQQLKKKCQRLTAELQ
                     DTKLHLEGQQVRNHELEKKQRRFDSELSQAHEETQREKLQREKLQREKDMLLAEAFSL
                     KQQMEEKDLDIAGFTQKVVSLEAELQDISSQESKDEASLAKVKKQLRDLEAKVKDQEE
                     ELDEQAGSIQMLEQAKLRLEMEMERMRQTHSKEMESRDEEVEEARQSCQKKLKQMEVQ
                     LEEEYEDKQKALREKRELESKLSTLSDQVNQRDFESEKRLRKDLKRTKALLADAQIML
                     DHLKNNAPSKREIAQLKNQLEESEFTCAAAVKARKAMEVEMEDLHLQIDDIAKAKTAL
                     EEQLSRLQREKNEIQNRLEEDQEDMNELMKKHKAAVAQASRDMAQMNDLQAQIEESNK
                     EKQELQEKLQALQSQVEFLEQSMVDKSLVSRQEAKIRELETRLEFEKTQVKRLENLAS
                     RLKETMEKLTEERDQRAAAENREKEQNKRLQRQLRDTKEEMSELARKEAEASRKKHEL
                     EMDLESLEAANQSLQADLKLAFKRIGDLQAAIEDEMESDENEDLINSEGDSDVDSELE
                     DRVDGVKSWLSKNKGPSKAPSDDGSLKSSSPTSHWKPLAPDPSDDEHDPVDSISRPRF
                     SHSYLSDSDTEAKLTETSA"
     repeat_region   68. .179
                     /note="KE repeat"
     misc_feature    635. .981
                     /gene="MysPDZ"
                     /note="PDZ domain"
     misc_feature    1272. .3558
                     /gene="MysPDZ"
                     /note="myosin head"
     misc_binding    1544. .1565
                     /gene="MysPDZ"
                     /bound_moiety="ATP/GTP binding site"
     misc_feature    3602. .3665
                     /gene="MysPDZ"
                     /note="IQ domain"
     misc_feature    3777. .5865
                     /gene="MysPDZ"
                     /note="coiled-coil"
     polyA_signal    7279. .7284
     polyA_site      7303
BASE COUNT     1768 a   1954 c   2232 g   1364 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 94;  Length 7318;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||
Db    288 CGTCAGGTGCAGGTCAG 272


RESULT  19
AC021767/c
LOCUS       AC021767    69900 bp    DNA             HTG       13-JUL-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-47D19 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC021767
VERSION     AC021767.2  GI:7144812
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 69900)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-47D19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69900)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6721397.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3201
                Center clone name: 47_D_19
            --------------.
            * NOTE: This record contains 79 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      787: contig of 787 bp in length
            *     788 887: gap of      100 bp
            *     888     1651: contig of 764 bp in length
            *    1652 1751: gap of      100 bp
            *    1752     2546: contig of 795 bp in length
            *    2547 2646: gap of      100 bp
            *    2647     3438: contig of 792 bp in length
            *    3439 3538: gap of      100 bp
            *    3539     4327: contig of 789 bp in length
            *    4328 4427: gap of      100 bp
            *    4428     5234: contig of 807 bp in length
            *    5235 5334: gap of      100 bp
            *    5335     6129: contig of 795 bp in length
            *    6130 6229: gap of      100 bp
            *    6230     7040: contig of 811 bp in length
            *    7041 7140: gap of      100 bp
            *    7141     7931: contig of 791 bp in length
            *    7932 8031: gap of      100 bp
            *    8032     8810: contig of 779 bp in length
            *    8811 8910: gap of      100 bp
```

```
*     8911     9696: contig of 786 bp in length
*     9697 9796: gap of      100 bp
*     9797    10596: contig of 800 bp in length
*    10597 10696: gap of      100 bp
*    10697    11493: contig of 797 bp in length
*    11494 11593: gap of      100 bp
*    11594    12399: contig of 806 bp in length
*    12400 12499: gap of      100 bp
*    12500    13296: contig of 797 bp in length
*    13297 13396: gap of      100 bp
*    13397    14152: contig of 756 bp in length
*    14153 14252: gap of      100 bp
*    14253    15034: contig of 782 bp in length
*    15035 15134: gap of      100 bp
*    15135    15922: contig of 788 bp in length
*    15923 16022: gap of      100 bp
*    16023    16795: contig of 773 bp in length
*    16796 16895: gap of      100 bp
*    16896    17698: contig of 803 bp in length
*    17699 17798: gap of      100 bp
*    17799    18568: contig of 770 bp in length
*    18569 18668: gap of      100 bp
*    18669    19457: contig of 789 bp in length
*    19458 19557: gap of      100 bp
*    19558    20342: contig of 785 bp in length
*    20343 20442: gap of      100 bp
*    20443    21239: contig of 797 bp in length
*    21240 21339: gap of      100 bp
*    21340    22122: contig of 783 bp in length
*    22123 22222: gap of      100 bp
*    22223    22995: contig of 773 bp in length
*    22996 23095: gap of      100 bp
*    23096    23836: contig of 741 bp in length
*    23837 23936: gap of      100 bp
*    23937    24720: contig of 784 bp in length
*    24721 24820: gap of      100 bp
*    24821    25590: contig of 770 bp in length
*    25591 25690: gap of      100 bp
*    25691    26480: contig of 790 bp in length
*    26481 26580: gap of      100 bp
*    26581    27368: contig of 788 bp in length
*    27369 27468: gap of      100 bp
*    27469    28258: contig of 790 bp in length
*    28259 28358: gap of      100 bp
*    28359    29134: contig of 776 bp in length
*    29135 29234: gap of      100 bp
*    29235    30016: contig of 782 bp in length
*    30017 30116: gap of      100 bp
*    30117    30897: contig of 781 bp in length
*    30898 30997: gap of      100 bp
*    30998    31770: contig of 773 bp in length
*    31771 31870: gap of      100 bp
*    31871    32662: contig of 792 bp in length
*    32663 32762: gap of      100 bp
*    32763    33558: contig of 796 bp in length
*    33559 33658: gap of      100 bp
*    33659    34448: contig of 790 bp in length
*    34449 34548: gap of      100 bp
*    34549    35346: contig of 798 bp in length
*    35347 35446: gap of      100 bp
*    35447    36251: contig of 805 bp in length
*    36252 36351: gap of      100 bp
*    36352    37136: contig of 785 bp in length
*    37137 37236: gap of      100 bp
*    37237    38021: contig of 785 bp in length
*    38022 38121: gap of      100 bp
*    38122    38908: contig of 787 bp in length
*    38909 39008: gap of      100 bp
*    39009    39800: contig of 792 bp in length
*    39801 39900: gap of      100 bp
*    39901    40683: contig of 783 bp in length
*    40684 40783: gap of      100 bp
*    40784    41551: contig of 768 bp in length
*    41552 41651: gap of      100 bp
*    41652    42409: contig of 758 bp in length
*    42410 42509: gap of      100 bp
*    42510    43302: contig of 793 bp in length
*    43303 43402: gap of      100 bp
*    43403    44204: contig of 802 bp in length
*    44205 44304: gap of      100 bp
*    44305    45097: contig of 793 bp in length
*    45098 45197: gap of      100 bp
*    45198    45976: contig of 779 bp in length
*    45977 46076: gap of      100 bp
*    46077    46856: contig of 780 bp in length
*    46857 46956: gap of      100 bp
*    46957    47735: contig of 779 bp in length
*    47736 47835: gap of      100 bp
*    47836    48609: contig of 774 bp in length
*    48610 48709: gap of      100 bp
*    48710    49519: contig of 810 bp in length
*    49520 49619: gap of      100 bp
*    49620    50405: contig of 786 bp in length
*    50406 50505: gap of      100 bp
*    50506    51282: contig of 777 bp in length
*    51283 51382: gap of      100 bp
*    51383    52152: contig of 770 bp in length
*    52153 52252: gap of      100 bp
*    52253    53044: contig of 792 bp in length
*    53045 53144: gap of      100 bp
*    53145    53941: contig of 797 bp in length
*    53942 54041: gap of      100 bp
*    54042    54836: contig of 795 bp in length
*    54837 54936: gap of      100 bp
*    54937    55730: contig of 794 bp in length
*    55731 55830: gap of      100 bp
*    55831    56616: contig of 786 bp in length
*    56617 56716: gap of      100 bp
*    56717    57498: contig of 782 bp in length
*    57499 57598: gap of      100 bp
*    57599    58381: contig of 783 bp in length
*    58382 58481: gap of      100 bp
*    58482    59250: contig of 769 bp in length
*    59251 59350: gap of      100 bp
*    59351    60139: contig of 789 bp in length
*    60140 60239: gap of      100 bp
*    60240    61030: contig of 791 bp in length
*    61031 61130: gap of      100 bp
*    61131    61917: contig of 787 bp in length
*    61918 62017: gap of      100 bp
*    62018    62811: contig of 794 bp in length

  Query Match              68.0%;  Score 17;  DB 66;  Length 69900;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||
Db   19825 CGTCAGGTGCAGGTCAG 19809


RESULT  20
AC069132
LOCUS       AC069132   151340 bp    DNA             HTG       17-JUN-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-556L8, *** SEQUENCING IN
            PROGRESS ***, 62 unordered pieces.
ACCESSION   AC069132
VERSION     AC069132.2  GI:8469032
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151340)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151340)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 12, 2000 this sequence version replaced gi:7924001.

            ---------- Genome Center ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 62 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1502: contig of 1502 bp in length
            *     1503     1602: gap of unknown length
            *     1603     2741: contig of 1139 bp in length
            *     2742     2841: gap of unknown length
            *     2842     4036: contig of 1195 bp in length
            *     4037     4136: gap of unknown length
            *     4137     5620: contig of 1484 bp in length
            *     5621     5720: gap of unknown length
            *     5721     7080: contig of 1360 bp in length
            *     7081     7180: gap of unknown length
            *     7181     8839: contig of 1659 bp in length
            *     8840     8939: gap of unknown length
            *     8940     9987: contig of 1048 bp in length
            *     9988    10087: gap of unknown length
            *    10088    11156: contig of 1069 bp in length
            *    11157    11256: gap of unknown length
            *    11257    12404: contig of 1148 bp in length
            *    12405    12504: gap of unknown length
            *    12505    13861: contig of 1357 bp in length
            *    13862    13961: gap of unknown length
            *    13962    15241: contig of 1280 bp in length
            *    15242    15341: gap of unknown length
            *    15342    17009: contig of 1668 bp in length
            *    17010    17109: gap of unknown length
            *    17110    18897: contig of 1788 bp in length
            *    18898    18997: gap of unknown length
            *    18998    20389: contig of 1392 bp in length
            *    20390    20489: gap of unknown length
            *    20490    22042: contig of 1553 bp in length
            *    22043    22142: gap of unknown length
            *    22143    24057: contig of 1915 bp in length
            *    24058    24157: gap of unknown length
            *    24158    26189: contig of 2032 bp in length
            *    26190    26289: gap of unknown length
            *    26290    27959: contig of 1670 bp in length
            *    27960    28059: gap of unknown length
            *    28060    30067: contig of 2008 bp in length
            *    30068    30167: gap of unknown length
            *    30168    32456: contig of 2289 bp in length
            *    32457    32556: gap of unknown length
            *    32557    34597: contig of 2041 bp in length
            *    34598    34697: gap of unknown length
            *    34698    36611: contig of 1914 bp in length
            *    36612    36711: gap of unknown length
            *    36712    38677: contig of 1966 bp in length
            *    38678    38777: gap of unknown length
            *    38778    40974: contig of 2197 bp in length
            *    40975    41074: gap of unknown length
            *    41075    42666: contig of 1592 bp in length
            *    42667    42766: gap of unknown length
            *    42767    44292: contig of 1526 bp in length
            *    44293    44392: gap of unknown length
            *    44393    46032: contig of 1640 bp in length
            *    46033    46132: gap of unknown length
            *    46133    48815: contig of 2683 bp in length
            *    48816    48915: gap of unknown length
            *    48916    51058: contig of 2143 bp in length
            *    51059    51158: gap of unknown length
            *    51159    53094: contig of 1936 bp in length
            *    53095    53194: gap of unknown length
            *    53195    54724: contig of 1530 bp in length
            *    54725    54824: gap of unknown length
            *    54825    57019: contig of 2195 bp in length
            *    57020    57119: gap of unknown length
            *    57120    59736: contig of 2617 bp in length
            *    59737    59836: gap of unknown length
            *    59837    61510: contig of 1674 bp in length
            *    61511    61610: gap of unknown length
            *    61611    63247: contig of 1637 bp in length
            *    63248    63347: gap of unknown length
            *    63348    65262: contig of 1915 bp in length
            *    65263    65362: gap of unknown length
            *    65363    67493: contig of 2131 bp in length
            *    67494    67593: gap of unknown length
            *    67594    70740: contig of 3147 bp in length
            *    70741    70840: gap of unknown length
            *    70841    72256: contig of 1416 bp in length
            *    72257    72356: gap of unknown length
            *    72357    75456: contig of 3100 bp in length
            *    75457    75556: gap of unknown length
            *    75557    77328: contig of 1772 bp in length
            *    77329    77428: gap of unknown length
            *    77429    79851: contig of 2423 bp in length
            *    79852    79951: gap of unknown length
            *    79952    82877: contig of 2926 bp in length
            *    82878    82977: gap of unknown length
            *    82978    84859: contig of 1882 bp in length
            *    84860    84959: gap of unknown length
            *    84960    87856: contig of 2897 bp in length
            *    87857    87956: gap of unknown length
            *    87957    91677: contig of 3721 bp in length
            *    91678    91777: gap of unknown length
            *    91778    94536: contig of 2759 bp in length
            *    94537    94636: gap of unknown length
            *    94637    97377: contig of 2741 bp in length
            *    97378    97477: gap of unknown length
            *    97478   100981: contig of 3504 bp in length
            *   100982   101081: gap of unknown length
            *   101082   103958: contig of 2877 bp in length
            *   103959   104058: gap of unknown length
            *   104059   106903: contig of 2845 bp in length
            *   106904   107003: gap of unknown length
            *   107004   109662: contig of 2659 bp in length
            *   109663   109762: gap of unknown length
            *   109763   113672: contig of 3910 bp in length
            *   113673   113772: gap of unknown length
            *   113773   118353: contig of 4581 bp in length
            *   118354   118453: gap of unknown length
            *   118454   121541: contig of 3088 bp in length
            *   121542   121641: gap of unknown length
            *   121642   124122: contig of 2481 bp in length
            *   124123   124222: gap of unknown length
            *   124223   128114: contig of 3892 bp in length
            *   128115   128214: gap of unknown length
            *   128215   131684: contig of 3470 bp in length
            *   131685   131784: gap of unknown length
            *   131785   135575: contig of 3791 bp in length
            *   135576   135675: gap of unknown length
            *   135676   140587: contig of 4912 bp in length
            *   140588   140687: gap of unknown length
            *   140688   145142: contig of 4455 bp in length
            *   145143   145242: gap of unknown length
            *   145243   151340: contig of 6098 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
```

```
                     /clone="RP11-556L8"
BASE COUNT     43058 a  29029 c  29780 g  43356 t   6117 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 74;  Length 151340;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||
Db   93019 CGTCAGGTGCAGGTCAG 93035


RESULT  21
AC068100
LOCUS       AC068100   155204 bp    DNA             HTG       16-MAR-2001
DEFINITION  Homo sapiens chromosome X clone RP11-812N12 map X, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC068100
VERSION     AC068100.4  GI:13357469
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155204)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome X, clone RP11-812N12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155204)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:12583851.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7131
                Center clone name: 812_N_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 6% of reads
                Sequencing vector: Plasmid; n/a; 94% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 146541 bases at least Q40
                Consensus quality: 150817 bases at least Q30
                Consensus quality: 152368 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 153204; sum-of-contigs
                Quality coverage: 4.5 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      518: contig of 518 bp in length
            *     519 618: gap of      100 bp
            *     619     1733: contig of 1115 bp in length
            *    1734 1833: gap of      100 bp
            *    1834     2717: contig of 884 bp in length
            *    2718 2817: gap of      100 bp
            *    2818     5343: contig of 2526 bp in length
            *    5344 5443: gap of      100 bp
            *    5444     8556: contig of 3113 bp in length
            *    8557 8656: gap of      100 bp
            *    8657    12409: contig of 3753 bp in length
            *   12410 12509: gap of      100 bp
            *   12510    16818: contig of 4309 bp in length
            *   16819 16918: gap of      100 bp
            *   16919    18984: contig of 2066 bp in length
            *   18985 19084: gap of      100 bp
            *   19085    41469: contig of 22385 bp in length
            *   41470 41569: gap of      100 bp
            *   41570    48545: contig of 6976 bp in length
            *   48546 48645: gap of      100 bp
            *   48646    55246: contig of 6601 bp in length
            *   55247 55346: gap of      100 bp
            *   55347    62502: contig of 7156 bp in length
            *   62503 62602: gap of      100 bp
            *   62603    70135: contig of 7533 bp in length
            *   70136 70235: gap of      100 bp
            *   70236    78809: contig of 8574 bp in length
            *   78810 78909: gap of      100 bp
            *   78910    88980: contig of 10071 bp in length
            *   88981 89080: gap of      100 bp
            *   89081   100598: contig of 11518 bp in length
            *  100599 100698: gap of      100 bp
            *  100699   115295: contig of 14597 bp in length
            *  115296 115395: gap of      100 bp
            *  115396   128681: contig of 13286 bp in length
            *  128682 128781: gap of      100 bp
            *  128782   140301: contig of 11520 bp in length
            *  140302 140401: gap of      100 bp
            *  140402   150306: contig of 9905 bp in length
            *  150307 150406: gap of      100 bp
            *  150407   155204: contig of 4798 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155204
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="RP11-812N12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .518
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    619. .1733
                     /note="assembly_fragment"
     misc_feature    1834. .2717
                     /note="assembly_fragment"
     misc_feature    2818. .5343
                     /note="assembly_fragment"
```

```
     misc_feature    5444. .8556
                     /note="assembly_fragment"
     misc_feature    8657. .12409
                     /note="assembly_fragment"
     misc_feature    12510. .16818
                     /note="assembly_fragment"
     misc_feature    16919. .18984
                     /note="assembly_fragment"
     misc_feature    19085. .41469
                     /note="assembly_fragment"
     misc_feature    41570. .48545
                     /note="assembly_fragment"
     misc_feature    48646. .55246
                     /note="assembly_fragment"
     misc_feature    55347. .62502
                     /note="assembly_fragment"
     misc_feature    62603. .70135
                     /note="assembly_fragment"
     misc_feature    70236. .78809
                     /note="assembly_fragment"
     misc_feature    78910. .88980
                     /note="assembly_fragment"
     misc_feature    89081. .100598
                     /note="assembly_fragment"
     misc_feature    100699. .115295
                     /note="assembly_fragment"
     misc_feature    115396. .128681
                     /note="assembly_fragment"
     misc_feature    128782. .140301
                     /note="assembly_fragment"
     misc_feature    140402. .150306
                     /note="assembly_fragment"
     misc_feature    150407. .155204
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    45793 a  31814 c  32000 g  43591 t   2006 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 73;  Length 155204;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db 119987 CGTCAGGTGCAGGTCAG 120003


RESULT  22
AC041048/c
LOCUS       AC041048   168360 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-759J14, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC041048
VERSION     AC041048.3  GI:7770701
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168360)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168360)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7622530.
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0759J14
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157456 bases at least Q40
            Consensus quality: 161145 bases at least Q30
            Consensus quality: 163068 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 166760; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.27 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1711: contig of 1711 bp in length
            *     1712     1811: gap of unknown length
            *     1812     6135: contig of 4324 bp in length
            *     6136     6235: gap of unknown length
            *     6236     8687: contig of 2452 bp in length
            *     8688     8787: gap of unknown length
            *     8788    11647: contig of 2860 bp in length
            *    11648    11747: gap of unknown length
            *    11748    13718: contig of 1971 bp in length
            *    13719    13818: gap of unknown length
            *    13819    18016: contig of 4198 bp in length
            *    18017    18116: gap of unknown length
            *    18117    21570: contig of 3454 bp in length
            *    21571    21670: gap of unknown length
            *    21671    24723: contig of 3053 bp in length
            *    24724    24823: gap of unknown length
            *    24824    32650: contig of 7827 bp in length
            *    32651    32750: gap of unknown length
            *    32751    46491: contig of 13741 bp in length
            *    46492    46591: gap of unknown length
            *    46592    60127: contig of 13536 bp in length
            *    60128    60227: gap of unknown length
            *    60228    74481: contig of 14254 bp in length
            *    74482    74581: gap of unknown length
            *    74582    90886: contig of 16305 bp in length
            *    90887    90986: gap of unknown length
            *    90987   104984: contig of 13998 bp in length
            *   104985   105084: gap of unknown length
            *   105085   127379: contig of 22295 bp in length
            *   127380   127479: gap of unknown length
            *   127480   147028: contig of 19549 bp in length
            *   147029   147128: gap of unknown length
            *   147129   168360: contig of 21232 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-759J14"
     misc_feature    1. .1711
                     /note="assembly_name:Contig6"
     misc_feature    1812. .6135
                     /note="assembly_name:Contig7"
     misc_feature    6236. .8687
                     /note="assembly_name:Contig8
```

```
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8788. .11647
                     /note="assembly_name:Contig9"
     misc_feature    11748. .13718
                     /note="assembly_name:Contig10"
     misc_feature    13819. .18016
                     /note="assembly_name:Contig11"
     misc_feature    18117. .21570
                     /note="assembly_name:Contig12"
     misc_feature    21671. .24723
                     /note="assembly_name:Contig13"
     misc_feature    24824. .32650
                     /note="assembly_name:Contig14"
     misc_feature    32751. .46491
                     /note="assembly_name:Contig15"
     misc_feature    46592. .60127
                     /note="assembly_name:Contig16"
     misc_feature    60228. .74481
                     /note="assembly_name:Contig17"
     misc_feature    74582. .90886
                     /note="assembly_name:Contig18"
     misc_feature    90987. .104984
                     /note="assembly_name:Contig19"
     misc_feature    105085. .127379
                     /note="assembly_name:Contig20"
     misc_feature    127480. .147028
                     /note="assembly_name:Contig21"
     misc_feature    147129. .168360
                     /note="assembly_name:Contig22"
BASE COUNT    48985 a  34329 c  34643 g  48785 t   1618 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 71;  Length 168360;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db 153451 CGTCAGGTGCAGGTCAG 153435


RESULT  23
AC060833/c
LOCUS       AC060833   176379 bp    DNA             HTG       17-AUG-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-556F3, WORKING DRAFT SEQUENCE,
            4 unordered pieces.
ACCESSION   AC060833
VERSION     AC060833.3  GI:9838272
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176379)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176379)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:8980022.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0556F03
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173730 bases at least Q40
            Consensus quality: 174579 bases at least Q30
            Consensus quality: 175046 bases at least Q20
            Insert size: 174000; agarose-fp
            Insert size: 177415; sum-of-contigs
            Quality coverage: 6.66 in Q20 bases; agarose-fp
            Quality coverage: 6.55 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1419: contig of 1419 bp in length
            *     1420     1519: gap of unknown length
            *     1520    10016: contig of 8497 bp in length
            *    10017    10116: gap of unknown length
            *    10117    79219: contig of 69103 bp in length
            *    79220    79319: gap of unknown length
            *    79320   176379: contig of 97060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176379
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-556F3"
     misc_feature    1. .1419
                     /note="assembly_name:Contig5"
     misc_feature    1520. .10016
                     /note="assembly_name:Contig6
                     clone_end:SP6
                     vector_side:right"
     misc_feature    10117. .79219
                     /note="assembly_name:Contig7"
     misc_feature    79320. .176379
                     /note="assembly_name:Contig8
                     clone_end:T7
                     vector_side:left"
BASE COUNT    49569 a  33984 c  35497 g  57025 t    304 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 72;  Length 176379;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db  81954 CGTCAGGTGCAGGTCAG 81938


RESULT  24
AC022781
LOCUS       AC022781   200729 bp    DNA             HTG       26-SEP-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC022781
VERSION     AC022781.5  GI:10305259
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200729)
```

```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-199H17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200729)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:9154653.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5913
                Center clone name: 199_H_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 180464 bases at least Q40
                Consensus quality: 190837 bases at least Q30
                Consensus quality: 194945 bases at least Q20
                Insert size: 206000; agarose-fp
                Insert size: 197129; sum-of-contigs
                Quality coverage: 3.9 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7021: contig of 7021 bp in length
            *     7022  7121: gap of      100 bp
            *     7122     8310: contig of 1189 bp in length
            *     8311  8410: gap of      100 bp
            *     8411     9411: contig of 1001 bp in length
            *     9412  9511: gap of      100 bp
            *     9512    11581: contig of 2070 bp in length
            *    11582 11681: gap of      100 bp
            *    11682    13342: contig of 1661 bp in length
            *    13343 13442: gap of      100 bp
            *    13443    16024: contig of 2582 bp in length
            *    16025 16124: gap of      100 bp
            *    16125    17440: contig of 1316 bp in length
            *    17441 17540: gap of      100 bp
            *    17541    18922: contig of 1382 bp in length
            *    18923 19022: gap of      100 bp
            *    19023    21353: contig of 2331 bp in length
            *    21354 21453: gap of      100 bp
            *    21454    23674: contig of 2221 bp in length
            *    23675 23774: gap of      100 bp
            *    23775    25868: contig of 2094 bp in length
            *    25869 25968: gap of      100 bp
            *    25969    27966: contig of 1998 bp in length
            *    27967 28066: gap of      100 bp
            *    28067    31051: contig of 2985 bp in length
            *    31052 31151: gap of      100 bp
            *    31152    33236: contig of 2085 bp in length
            *    33237 33336: gap of      100 bp
            *    33337    36285: contig of 2949 bp in length
            *    36286 36385: gap of      100 bp
            *    36386    39476: contig of 3091 bp in length
            *    39477 39576: gap of      100 bp
            *    39577    42584: contig of 3008 bp in length
            *    42585 42684: gap of      100 bp
            *    42685    46164: contig of 3480 bp in length
            *    46165 46264: gap of      100 bp
            *    46265    49973: contig of 3709 bp in length
            *    49974 50073: gap of      100 bp
            *    50074    55389: contig of 5316 bp in length
            *    55390 55489: gap of      100 bp
            *    55490    59283: contig of 3794 bp in length
            *    59284 59383: gap of      100 bp
            *    59384    62877: contig of 3494 bp in length
            *    62878 62977: gap of      100 bp
            *    62978    68182: contig of 5205 bp in length
            *    68183 68282: gap of      100 bp
            *    68283    73862: contig of 5580 bp in length
            *    73863 73962: gap of      100 bp
            *    73963    83038: contig of 9076 bp in length
            *    83039 83138: gap of      100 bp
            *    83139    87822: contig of 4684 bp in length
            *    87823 87922: gap of      100 bp
            *    87923    92086: contig of 4164 bp in length
            *    92087 92186: gap of      100 bp
            *    92187    97007: contig of 4821 bp in length
            *    97008 97107: gap of      100 bp
            *    97108   123018: contig of 25911 bp in length
            *   123019 123118: gap of      100 bp
            *   123119   129307: contig of 6189 bp in length
            *   129308 129407: gap of      100 bp
            *   129408   137163: contig of 7756 bp in length
            *   137164 137263: gap of      100 bp
            *   137264   147510: contig of 10247 bp in length
            *   147511 147610: gap of      100 bp
            *   147611   155502: contig of 7892 bp in length
            *   155503 155602: gap of      100 bp
            *   155603   168939: contig of 13337 bp in length
            *   168940 169039: gap of      100 bp
            *   169040   181336: contig of 12297 bp in length
            *   181337 181436: gap of      100 bp
            *   181437   195395: contig of 13959 bp in length
            *   195396 195495: gap of      100 bp
            *   195496   200729: contig of 5234 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200729
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /map="11"
                     /clone="RP23-199H17"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .7021
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7122. .8310
                     /note="assembly_fragment"
     misc_feature    8411. .9411
                     /note="assembly_fragment"
     misc_feature    9512. .11581
```

```
                     /note="assembly_fragment"
     misc_feature    11682. .13342
                     /note="assembly_fragment"
     misc_feature    13443. .16024
                     /note="assembly_fragment"
     misc_feature    16125. .17440
                     /note="assembly_fragment"
     misc_feature    17541. .18922
                     /note="assembly_fragment"
     misc_feature    19023. .21353
                     /note="assembly_fragment"
     misc_feature    21454. .23674
                     /note="assembly_fragment"
     misc_feature    23775. .25868
                     /note="assembly_fragment"
     misc_feature    25969. .27966
                     /note="assembly_fragment"
     misc_feature    28067. .31051
                     /note="assembly_fragment"
     misc_feature    31152. .33236
                     /note="assembly_fragment"
     misc_feature    33337. .36285
                     /note="assembly_fragment"
     misc_feature    36386. .39476
                     /note="assembly_fragment"
     misc_feature    39577. .42584
                     /note="assembly_fragment"
     misc_feature    42685. .46164
                     /note="assembly_fragment"
     misc_feature    46265. .49973
                     /note="assembly_fragment"
     misc_feature    50074. .55389
                     /note="assembly_fragment"
     misc_feature    55490. .59283
                     /note="assembly_fragment"
     misc_feature    59384. .62877
                     /note="assembly_fragment"
     misc_feature    62978. .68182
                     /note="assembly_fragment"
     misc_feature    68283. .73862
                     /note="assembly_fragment"
     misc_feature    73963. .83038
                     /note="assembly_fragment"

  Query Match             68.0%;  Score 17;  DB 67;  Length 200729;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           |||||||||||||||||
Db   69699 CGTCAGGTGCAGGTCAG 69715


RESULT  25
CNS078T3/c
LOCUS       CNS078T3     1044 bp    DNA             STS       12-JAN-2001
DEFINITION  T7 end of clone BB0AA023D11 of library BB0AA from strain CBS 4732
            of Pichia angusta, sequence tagged site.
ACCESSION   AL434413
VERSION     AL434413.1  GI:12217827
KEYWORDS    STS.
SOURCE      Pichia angusta.
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1044)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
   PUBMED   11152888
REFERENCE   2  (bases 1 to 1044)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 1044)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1044
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA023D11"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    complement(<1. .>999)
                     /note="similar to Saccharomyces cerevisiae ORF YGL026c [
                     TRP5 ; tryptophan synthase ]"
                     /evidence=not_experimental
BASE COUNT      245 a    311 c    228 g    257 t      3 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 53;  Length 1044;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacg 22
           ||||||||||||||||
Db     546 GGTGCAGGTCAGCACG 531


RESULT  26
AF237379
LOCUS       AF237379     8326 bp    DNA             VRT       13-NOV-2000
DEFINITION  Xenopus laevis neurofilament protein (NF-M(2)) gene, complete cds.
ACCESSION   AF237379
VERSION     AF237379.1  GI:11139431
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 8326)
  AUTHORS   Roosa,J.R., Gervasi,C. and Szaro,B.G.
  TITLE     Structure, biological activity of the upstream regulatory sequence,
            and conserved domains of a middle molecular mass neurofilament gene
            of Xenopus laevis
  JOURNAL   Brain Res. Mol. Brain Res. 82 (1-2), 35-51 (2000)
   PUBMED   11042356
REFERENCE   2  (bases 1 to 8326)
  AUTHORS   Roosa,J.R. and Szaro,B.G.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (18-FEB-2000) Biological Sciences, University at Albany,
            State University of New York, 1400 Washington Avenue, Albany, NY
            12222, USA
FEATURES             Location/Qualifiers
     source          1. .8326
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     TATA_signal     1498. .1503
     5'UTR           1524. .1563
     exon            1524. .2598
                     /number=1
     mRNA            join(1524. .2598,3533. .3657,6507. .>8326)
                     /product="neurofilament protein"
     gene            1564. .>8326
                     /gene="NF-M(2)"
     CDS             join(1564. .2598,3533. .3657,6507. .8088)
                     /gene="NF-M(2)"
                     /codon_start=1
                     /product="neurofilament protein"
                     /protein_id="AAG31697.1"
                     /db_xref="GI:11139432"
                     /translation="MSYSLETIGNPSYRRMAETRTSYSRTSSTPSSGFRSQSWSRSSP
                     TTISSYKRTNLAAPRVYSSNDSLDFSQSSAFNGDLKQVRFSEKEQLQGLNDRFVGYIE
                     KVHFLEQQNKEIEAEIQALRQKQSGYSQLGEVYEQEMRELRVNLELLGHDKAQIVLDT
                     EHLEDDLQRLKERLEEEARIRDETEAAIRAMKKDMEDTSLVKVELDKKVQSLQEEVAF
                     IKNNHEEEVGELLAQIQASSITVERKDFQKTDLASALREIRSQLEGHSSQNMLQTEEV
                     FRNRYAKLTEAAEQNKEAIRSAKDDIAEYRRQLHSKTVELESVRGTKESLERQLNDIE
                     ERHGHDLTSYQETIQQLDNELRGTKWEMSRHLREYQDLLNVKMALDIEIAAYRKLLEG
                     EETRFSSFSGVPTGPSAIHRPTPISSTRVQQARAEPPRVKVQHKFVEEIIEETKVEDD
                     RSDMEAALAAMADEFALGFEEGGEEEEQGEETQEGEEEAEEEIVAAVESSVQAAAPGE
                     EAEGEEEEEEKGEEEAEAEEEGEKEEEEKEEEGEDEGEKVDEEEEGGEQGEEEDVKAE
                     EEEEKEEGEEDAEVEETKAEEGEEEGEKGEDEAEKEEGGEEEEKEEVAADEEGGEGEE
                     EGGEGEKDEDKGEEEKEGEAEAEEAEGGEEETTEEAQAEEVVEETMTETKIVREKAEV
                     QETEAEEEETQDAEEQKDEGEDEETGEGKETEEKEEGGEEKEDADEGTEQQEEEKAEE
                     EKGEIDEGETKEGAEEDAGEDKQAKSPEEKSSEETISVSKAASVGSEKESAKQEEEKS
                     TQRDDGKEDTAVNGDVEEGEVKEESEDKAVVTNGLDDSPSEDKESSKGQQTVVITKKI
                     ETTTSEGEDGTKYITKSVTVTQKVEEHEETVQEKLVSTKKVEKVTSHAVVKEVTEDK"
     intron          2599. .3532
                     /gene="NF-M(2)"
                     /number=1
     exon            3533. .3657
                     /gene="NF-M(2)"
                     /number=2
     intron          3659. .6507
                     /gene="NF-M(2)"
                     /number=2
     exon            6507. .>8326
                     /gene="NF-M(2)"
                     /number=3
     3'UTR           8089. .>8326
                     /gene="NF-M(2)"
BASE COUNT     2631 a   1560 c   2161 g   1974 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 8;  Length 8326;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db    3060 CAGGTGCAGGTCAGCA 3075


RESULT  27
AE004050
LOCUS       AE004050    10536 bp    DNA             BCT       06-JUL-2000
DEFINITION  Xylella fastidiosa, section 196 of 229 of the complete genome.
ACCESSION   AE004050 AE003849
VERSION     AE004050.1  GI:9107594
KEYWORDS    .
SOURCE      Xylella fastidiosa.
  ORGANISM  Xylella fastidiosa
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10536)
  AUTHORS   Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
            Terenzi,M.F., Truffi,D., Tsai,S.M., Tsuhako,M.H., Vallada,H., Van
            Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
            Meidanis,J. and Setubai,J.C.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis, Sao Paulo, Brazil
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
REFERENCE   2  (bases 1 to 10536)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
  REMARK
FEATURES             Location/Qualifiers
     source          1. .10536
                     /organism="Xylella fastidiosa"
                     /db_xref="taxon:2371"
     gene            1296. .2339
                     /gene="XF2409"
     CDS             1296. .2339
                     /gene="XF2409"
                     /note="similar to SP|P39369 (percent identity: 47 %/query
                     alignment coverage:  96.8 %/subject alignment coverage:
                     99.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder/Start codon shift: 2523"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF85208.1"
                     /db_xref="GI:9107595"
                     /translation="MPALGYLHVDGLCQKSSGGSRYNLYEAETIAAWLAEHHESLQAQ
                     YGEPLHRIVGILTPFETQARAISQACRDVGIEVGHEKDGITVGTVHALQGAERPVVIF
                     SAVYSKHADGGFIDQRASMLNVAVSRAKNTFLVFGDMDVFTAAPKSRPRGLLAHYLFK
                     DPSNALCFQPLVRKDLQQVSTAVEVLQDAAEHDAFLLQALNKVQREIHIVSPWINKDR
                     IQDIGAFKAMQEAVERQVQVTVYTDQDLNTDDKKDTKKIAEVLQAARALRGVGIEVHF
                     VDRVHSKMVIGDDEVFCVGSFNWFSANRSAMYAKHETSLVYRGRGLADERQTRLNSLR
                     QRITDDPQQVGQA"
     gene            2311. .2484
                     /gene="XF2410"
     CDS             2311. .2484
                     /gene="XF2410"
                     /note="hypothetical protein; identified by sequence
```

```
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85209.1"
                     /db_xref="GI:9107596"
                     /translation="MILSKSARRDAVLRQRDVDTGWRQAMHAAAEGRDMVSPAVALRA
                     VALLKLCCCLPLG"
     gene            complement(3181. .5514)
                     /gene="XF2411"
     CDS             complement(3181. .5514)
                     /gene="XF2411"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85210.1"
                     /db_xref="GI:9107597"
                     /translation="MLKKLIKVNHVGLLRNGVPPTTPTPFGPVTLFYAENGRGKTTVA
                     NILRAVTHGKDKDEVLQKARTIDTDDPPCIELLWNDGNTDIKVTLHHGHWTGTPPDIL
                     IFDATFVDDNVYSGQEVHPAHRQKFLQFVLGAESVKLEREIVRLAKEIADEHRNIKER
                     EQRIRAYSKNMPIADFIRLIPVTDAEAKIAALRTQREAVRNAPALLQRHAPQPLPHIE
                     CNIDAFFSIIQSTFHDIRADAEEIVQAHFQRHQQPPGIEAWVMQGREFNAEQGCPFCG
                     QEVTATALIKAYETYFNTAFKAFMKNVATLGSGVATRFPEACLEKIADRIQGNDACVA
                     AWKDQLDLSSPFFDNPRATATFTRLRTVALALAQSKQHRPLDCVGTTQDKADMVRLLE
                     ELRAQMDAYNAAVVEINRRITNFRKTLAENTPAALDASIATLEIGIVRQSTEVVQAIT
                     VYQAAKAKKEQLEREKKNMRAALDAGLPDLLSEYASKINQFLRDFGAAFSIKKLQQSM
                     QGGTMRASYVLQLRGKEVALGRRTDSDPGFHNVLSEGDKRTLALAFFLARLYATPDAL
                     VGKSVVLDDPMCSFDMTRRNRTMESIAALVNQGVQVVVLSHDAYFLRDLRDLLAGPRY
                     NKVSVNVHHIKRTKNNDSQIVSDADLDTICQSPYMLRYAQVVTFVSGTYEGTLQEVAS
                     ALRPLVEGFLKHRFAPPLLRQDLSLGQMISDIRKATDDSPLVLAKPYVDTLEKLNAFA
                     VQSHHDDSKSLSPMNDGQLRQYARMALELIYGGSLPH"
     gene            complement(5615. .5800)
                     /gene="XF2412"
     CDS             complement(5615. .5800)
                     /gene="XF2412"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85211.1"
                     /db_xref="GI:9107598"
                     /translation="MIFGATQTNADRRRDSGTRHAVCTGWARSRLPPDPQVRHNAHLP
                     KPHVKVISSIPITGTML"
     gene            6084. .6512
                     /gene="XF2413"
     CDS             6084. .6512
                     /gene="XF2413"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85212.1"
                     /db_xref="GI:9107599"
                     /translation="MVGGDVPTADTRAAAGGACEPSIKNIALSRGVHVLLLCAGMRLW
                     PGQCCLGQNGPMKGSVVKSSVVLGVVLRCGSQCQQVEAAARGFAPRTFGKYCMSALWL
                     YRRWCSERHCLSCRTMEGMGATSGDAGPRRGRGWLGKGRG"
     gene            complement(6509. .8314)
                     /gene="XF2414"
     CDS             complement(6509. .8314)
                     /gene="XF2414"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85213.1"
                     /db_xref="GI:9107600"
                     /translation="MPPKRHQVHDAADTDKENLTVTTEHCFIGVCFIPLDGSGQTFFG
                     FSEFLAGLALMVLAWTIADVRYRFRIQVAPLPLLRMSFYVVAAIGLLTLLTDLWRAEQ
                     WLVPRGILLTPAVWQALLGGAFLLTILTWAWLGFIRPPTYGQYNAARFQQTLYRIILN
                     GSPTELAVIADELVHSAEALVQHATNRYSLQNGERIECEEEEQEDLPQVNRDANALLL
                     LIAEKRLCRAIVAESPRTAFALFQAIANTKKYGIPIRVFAQNIFHEALANKDSFLYHE
                     TNGCALGLIGDHKPFTQAMFANYEMVETIGTLFNHDRDARNKWDFVQWLAYCSVVLMT
                     LRSYAKQGYRGYSFVLSQAMGYIASAASDLSTLNEPSSAADSDVLRRLALVIQFIKDA
                     VRILDQEGVPEHVCICVRAGYDSSRAIFCDRIANMIFTVISLASEVRLPQLEWKIQHN
                     SVWSNLFHYGAFKRPAGQVVAFKLRRLLYNQIADMQRFPNVERAKILRFCLNVLALAV
                     IRGNDYKSDRRFHHRVLHRAVLAWTKKHFVWLHSHHPRLAESCLGNHITYDAEHLRLA
                     YTHPAKGPIGKVSYFALNPASPPPEPVAASAVPNA"
     tRNA            complement(8365. .8440)
                     /gene="tRNA-Thr-2"
                     /note="found by tRNAscan"
                     /product="tRNA-Thr"
     gene            complement(8365. .8440)
                     /gene="tRNA-Thr-2"
     gene            complement(8457. .8648)
                     /gene="XF2415"
     CDS             complement(8457. .8648)
                     /gene="XF2415"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85214.1"
                     /db_xref="GI:9107601"
                     /translation="MPVPRAVDPPQPPILKSHLCATVAHLLQSIPSPDPALILMRRAP
                     EAHTKQSSALSRRPSTAMM"
     gene            complement(8921. .9871)
                     /gene="XF2416"
     CDS             complement(8921. .9871)
                     /gene="XF2416"
                     /note="similar to GI|5915671 (percent identity: 64 %/query
                     alignment coverage:  99.4 %/subject alignment coverage:
                     99.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="drug tolerance protein"
                     /protein_id="AAF85215.1"
                     /db_xref="GI:9107602"
                     /translation="MKVLLANPRGFCAGVDRAIEIVKRTIDMLGTPIYVRHEVVHNRF
                     VVDDLKQRGAIFVEELHQVPDGATVIFSAHGVSQAVRRQAAQRGLKVFDATCPLVTKV
                     HLDVARHCRTGRDMILIGHAGHPEVEGTMGQWDQERGTGRIYLVENIDDVATLHVAQP
                     HHLAYTTQTTLSVDDTRNIIDALRQRFPTIQGPKNNDICYATQNRQDAVRELARECDL
                     VLVVGSPNSSNSNRLSELAQREGVASYLIDSAAEIDPAWVIDKHHIGVTAGASAPQVL
                     VDGVLARLYELGATSVSEHSGKPESMVFALPKALRLQLVD"

  Query Match             64.0%;  Score 16;  DB 1;  Length 10536;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtca 17
           ||||||||||||||||
Db    1995 CGTCAGGTGCAGGTCA 2010


RESULT  28
AE005241
LOCUS       AE005241    12242 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 65
            of 155.
ACCESSION   AE005241 AE005174
VERSION     AE005241.1  GI:12513509
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 12242)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
```

```
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 12242)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .12242
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            complement(146. .952)
                     /gene="rna"
                     /note="Z0755"
     CDS             complement(146. .952)
                     /gene="rna"
                     /function="enzyme; Macromolecule degradation: Degradation
                     of RNA"
                     /note="Residues 1 to 268 of 268 are 99.25 pct identical to
                     residues 1 to 268 of 268 from Escherichia coli K-12 Strain
                     MG1655: B0611"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase I, cleaves phosphodiester bond between any
                     two nucleotides"
                     /protein_id="AAG54946.1"
                     /db_xref="GI:12513510"
                     /translation="MKAFWRNAALLAVSLLPFSSANAVALQAKQYGDFDRYVLALSWQ
                     TGFCQSQHDRNRNERDECRLQTETTNKADFLTVHGLWPXLPKSVAARGVDERRWMRFG
                     CATRPIPNLPEARASRMCSSPETGLSLETAAKLSEVMPGAGGRSCLERYEYAKHGACF
                     GFDPDAYFGTMVRLNQEIKESEAGKFLADNYGKTVSRRDFDAAFAKSWGKENVKAVKL
                     TCQGNPAYLTEIQVSIKADAINAPLSANSFLPQPHPGNCGKTFVIDKAGY"
     gene            complement(1066. .2529)
                     /gene="ybdS"
                     /note="Z0756"
     CDS             complement(1066. .2529)
                     /gene="ybdS"
                     /function="putative membrane; Not classified"
                     /note="Residues 1 to 487 of 487 are 100.00 pct identical
                     to residues 1 to 487 of 487 from Escherichia coli K-12
                     Strain MG1655: B0612"
                     /codon_start=1
                     /transl_table=11
                     /product="putative a membrane protein"
                     /protein_id="AAG54947.1"
                     /db_xref="GI:12513511"
                     /translation="MSLAKDNIWKLLAPLVVMGVMFLIPVPDGMPPQAWHYFAVFVAM
                     IVGMILEPIPATAISFIAVTICVIGSNYLLFDAKELADPAFNAQKQALKWGLAGFSST
                     TVWLVFGAFIFALGYEVSGLGRRIALFLVKFMGKRTLTLGYAIVIIDILLAPFTPSNT
                     ARTGGTVFPVIKNLPPLFKSFPNDPSARRIGGYLMWMMVISTSLSSSMFVTGAAPNVL
                     GLEFVSKIAGIQISWLQWFLCFLPVGVILLIIAPWLSYVLYKPEITHSEEVATWAGDE
                     LKTMGALTRREWTLIGLVLLSLGLWVFGSEVINATAVGLLAVSLMLALHVVPWKDITR
                     YNSAWNTLVNLATLVVMANGLTRSGFIDWFAGTMSTHLEGFSPNATVIVLVLVFYFAH
                     YLFASLSAHTATMLPVILAVGKXIPGVPMEQLCILLVLSIGIMGCLTPYATGPGVIIY
                     GCGYVKSKDYWRLGAIFGVIYISMLLLVGWPILAMWN"
     gene            complement(2580. .3458)
                     /gene="citG"
                     /note="Z0757"
     CDS             complement(2580. .3458)
                     /gene="citG"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 292 of 292 are 98.97 pct identical to
                     residues 1 to 292 of 292 from Escherichia coli K-12 Strain
                     MG1655: B0613"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG54948.1"
                     /db_xref="GI:12513512"
                     /translation="MSMPATLTKTTKLATSLIDEYALLGWRAMLTEVNLSPKPGLVDR
                     INCGAHKDMALEDFHRSALAIQGWLPRFIEFGACSAEMAPEAVLHGLRPIGMACEGDM
                     FRATAGVNTHKGSIFSLGLLCAAIGRLLQLNQLLTPTTVCSTAASFCRGLTDRELRTN
                     NSQLTAGQRLYQQLGLTGARGEAEAGYPLVINHALPHYLTLLDQGLDPELALLDTLLL
                     LMAINGDTNVASRGGEGGLRWLQREAQTLLQKGGIRTPADLDYLRQFDRECIERNLSP
                     GGSADLLILTWFLAQI"
     gene            complement(3433. .3984)
                     /gene="ybdU"
                     /note="Z0758"
     CDS             complement(3433. .3984)
                     /gene="ybdU"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 183 of 183 are 99.45 pct identical to
                     residues 1 to 183 of 183 from Escherichia coli K-12 Strain
                     MG1655: B0614"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG54949.1"
                     /db_xref="GI:12513513"
                     /translation="MHLLPELASHHAVSIPELLVSRDERQARQHVWLKRHPVPLVSFT
                     VVAPGPIKDSEVTRRIFNHGVTALRALAAKQGWQIQEQAALVSASGPEGMLSIAAPAR
                     DLKLATIELEHXHPLGRLWDIDVLTPEGEILSRRDYSLPPRRCLLCEQSAAVCARGKT
                     HQLTDLLNRMEALLNDVDACNVN"
     gene            complement(3988. .5520)
                     /gene="citF"
                     /note="Z0759"
     CDS             complement(3988. .5520)
                     /gene="citF"
                     /function="enzyme; Central intermediary metabolism: Pool,
                     multipurpose conversions of intermed. met'm"
                     /note="Residues 1 to 510 of 510 are 99.41 pct identical to
                     residues 1 to 510 of 510 from Escherichia coli K-12 Strain
                     MG1655: B0615"
                     /codon_start=1
                     /transl_table=11
                     /product="citrate lyase alpha chain"
                     /protein_id="AAG54950.1"
                     /db_xref="GI:12513514"
                     /translation="MTQKIEQSQRQERVAAWNRRAECDLAAFQNSPKQTYQAEKARDR
                     KLCANLEEAIRRSGLQDGMTVSFHHAFRGGDLTVNMVMDVIAKMGFKNLILASSSLSD
                     CHAPLVEHIRQGVVTRIYTSGLRGPLAEEISRGLLAEPVQIHSHGGRVHLVQSGELNI
                     DVAFLGVPSCDEFGNANGYTGKACCGSLGYAMVDADNAKQVVMLTEELLPYPHNPASI
                     EQDQVDLIVKVDRVGDAAKIGAGATRMTTNPRELLIARSAADVIVNSGYFKEGFSMQT
                     GTGGASLAVTLFLEDKMRSRDIRADFALGGITATMVDLHEKGLIRKLLDVQSFDSHAA
                     QSLARNPNHIEISANQYANWGSKGASVDRLDVVVLSALEIDTQFNVNVLTGSDGVLRG
                     ASGGHCDTAIASALSIIVAPLVRGRIPTLVDNVLTCITPGSSVDILVTDHGIAVNPAR
                     PELAERLQEAGIKVVSIEWLRERARLLTGEPQPIEFTDRVVAVVRYRDGSVIDVVHQV
                     KE"
     gene            complement(5531. .6454)
                     /gene="citE"
                     /note="Z0760"
     CDS             complement(5531. .6454)
                     /gene="citE"
                     /function="enzyme; Central intermediary metabolism: Pool,
                     multipurpose conversions of intermed. met'm"
                     /note="Residues 1 to 307 of 307 are 99.34 pct identical to
                     residues 1 to 307 of 307 from Escherichia coli K-12 Strain
                     MG1655: B0616"
                     /codon_start=1
                     /transl_table=11
                     /product="citrate lyase beta chain (acyl lyase subunit)"
                     /protein_id="AAG54951.1"
                     /db_xref="GI:12513515"
```

```
                     /translation="MGGLPMISASLQQRKTRTRRSMLFVPGANAAMVSNSFIYPADAL
                     MFDLEESVALREKDTARRMVYHALQHPLYRDIETIVRVNALDSEWGVNDLEAVVRGGA
                     DVVRLPKTDTAQDVLDIEKEILRIEKACGREPGSTGLLAAIESPLGITRAVEIAHASE
                     RLIGIALGAEDYVRNLRTERSPEGTELLFARCSILQAARSAGIQAFDTVYSDANNEAG
                     FLQEAAHIKQLGFDGKSLINPRQIDLLHNLYAPTQKEVDHARRVVEAAEAAAREGLGV
                     VSLNGKMVDGPVIDRARLVXSRAELSGIREE"
     gene            complement(6436. .6732)
                     /gene="citD"
                     /note="Z0761"
     CDS             complement(6436. .6732)
                     /gene="citD"
                     /function="enzyme; Central intermediary metabolism: Pool,
                     multipurpose conversions of intermed. met'm"
                     /note="Residues 1 to 98 of 98 are 100.00 pct identical to
                     residues 1 to 98 of 98 from Escherichia coli K-12 Strain
                     MG1655: B0617"
                     /codon_start=1
                     /transl_table=11
                     /product="citrate lyase acyl carrier protein (gamma
                     chain)"
                     /protein_id="AAG54952.1"
                     /db_xref="GI:12513516"
                     /translation="MKINQPAVAGTLESGDVMIRIAPLDTQDIDLQINSSVEKQFGDA
                     IRTTILDVLARYNVRGVQLNVDDKGALDCILRARLEALLARASGIPALPWEDCQ"
     gene            complement(6747. .7892)
                     /gene="citC"
                     /note="Z0762"
     CDS             complement(6747. .7892)
                     /gene="citC"
                     /function="enzyme; Central intermediary metabolism: Pool,
                     multipurpose conversions of intermed. met'm"
                     /note="Residues 1 to 381 of 381 are 99.21 pct identical to
                     residues 1 to 381 of 381 from Escherichia coli K-12 Strain
                     MG1655: B0618"
                     /codon_start=1
                     /transl_table=11
                     /product="citrate lyase synthetase (citrate (pro-3S)-lyase
                     ligase"
                     /protein_id="AAG54953.1"
                     /db_xref="GI:12513517"

  Query Match             64.0%;  Score 16;  DB 1;  Length 12242;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcacgtt 24
           ||||||||||||||||
Db    4236 TGCAGGTCAGCACGTT 4251


RESULT  29
MLCB1450
LOCUS       MLCB1450    38065 bp    DNA             BCT       27-AUG-1999
DEFINITION  Mycobacterium leprae cosmid B1450.
ACCESSION   AL035159
VERSION     AL035159.1  GI:4154036
KEYWORDS    60 kD chaperonin; CDP-diacylglycerol--serine
            O-phosphatidyltransferase; cpn60; glutamine-binding protein;
            groEL-2; hydrolase; oxidoreductase; phosphomethylpyrimidine kinase;
            pseudogene; pssA; RLEP1; serine/threonine protein kinase; thiamine
            biosynthesis protein ThiG; thiamine-phosphate pyrophosphorylase;
            thiD; thiE; thiG.
SOURCE      Mycobacterium leprae.
  ORGANISM  Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 38065)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE     Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2  (bases 1 to 38065)
  AUTHORS   Brown,D. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38065)
  AUTHORS   James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1998) Mycobacterium leprae sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
            Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,
            Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            France Requests for cosmids should be sent to Karin Eiglmeier
            (kei@pasteur.fr)
COMMENT     Notes:
            The Sanger Centre is funded to complete the sequence of M. leprae
            by the Heiser Program for Research in Leprosy and Tuberculosis of
            The New York Community Trust.
            Work in Paris is supported by the Heiser Trust, the Association
            Francaise Raoul Follereau and the Groupement de Recherches et des
            Etudes des Genomes (GIP-GREG).
            Details of M. leprae sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/)
            CDS are numbered using the following system eg MLCB33.01c. ML (M.
            leprae), cB33 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. All CDS
            over 100 codons have been analysed.  Gene prediction is based on
            positional base preference in codons especially where there is an
            increase in the observed/expected third position G + C.  CAUTION:
            We may not have predicted the correct initiation codon.  Where
            possible we choose an initiation codon (atg, gtg, or ttg) which is
            preceded by an upstream ribosome binding site sequence (optimally
            5-13bp before the initiation codon).  If this cannot be identified
            we choose the most upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
FEATURES             Location/Qualifiers
     source          1. .38065
                     /organism="Mycobacterium leprae"
                     /db_xref="taxon:1769"
                     /clone="cosmid B1450"
     gene            complement(1. .284)
                     /gene="MLCB1450.01c"
     CDS             complement(<1. .284)
                     /gene="MLCB1450.01c"
                     /note="MLCB1450.01c, probable transcriptional regulator,
                     partial CDS, len: >94 aa; similar to e.g. TR:Q50887
                     (EMBL:Z56280) Myxococcus xanthus transcription factor CARD
                     (316 aa), fasta scores; opt: 190 z-score: 372.9 E():
                     1.8e-13, 28.6% identity in 98 aa overlap. Equivalent to
                     M.tuberculosis Rv3583c (97.9% identity in 94 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="CAA22685.1"
                     /db_xref="GI:4154037"
                     /db_xref="SPTREMBL:Q9ZBM8"
                     /translation="MIFKVGDTVVYPHHGAALVEAIETRTINGEQKEYLVLKVAQGDL
                     TVRVPAENAEYVGVRDVVGQEGLDQVFQVLRAPHTEEPTNWSRRYKANLE"
     gene            654. .1205
                     /gene="MLCB1450.02"
     CDS             654. .1205
                     /gene="MLCB1450.02"
                     /note="MLCB1450.02, probable lipoprotein, len: 183 aa;
```

```
                Equivalent to M.tuberculosis Rv3584 (MTV024.02), 63.4%
                identity in 175 aa overlap. Contains N-terminal signal
                sequence with appropriately positioned PS00013 Prokaryotic
                membrane lipoprotein lipid attachment site"
                /codon_start=1
                /transl_table=11
                /product="putative lipoprotein"
                /protein_id="CAA22686.1"
                /db_xref="GI:4154038"
                /db_xref="SPTREMBL:Q9ZBM7"
                /translation="MSRFKISLPALATRVAVLGFLTLMASVLGGCGAGQISQTATQEP
                AVNGNRVTLNNLALRDIRIQAAQTGDFLQSGRTVDLMLVAINNSPYVTDRLVSITSDI
                GTVALNGYTQLPTNGMLFIGTSEGQRIKPPPLQSNNIAKAIVTLAKPITNGLTYNFTF
                NFEKAGQANVAVPVSAGLAPRQT"
misc_feature    714. .746
                /gene="MLCB1450.02"
                /note="PS00013 Prokaryotic membrane lipoprotein lipid
                attachment site"
gene            1428. .2463
                /gene="radA"
CDS             1428. .2463
                /gene="radA"
                /note="MLCB1450.03, probable radA pseudogene, len: 1033
                bp, Equivalent to M.tuberculosis Rv3585, (MTV024.03),
                probable DNA repair protein RADA homolog"
                /codon_start=1
                /pseudo
                /transl_table=11
                /product="putative radA pseudogene"
                /db_xref="PID:e1370359"
misc_feature    complement(1723. .5324)
                /note="identical to EMBL:MLANTM M14341 M.leprae 65 kd
                antigen, from 12 to 3613"
repeat_unit     complement(2559. .3208)
                /note="RLEP1"
gene            2832. .3170
                /gene="MLCB1450.04"
CDS             2832. .3170
                /gene="MLCB1450.04"
                /note="MLCB1450.04, CDS in RLEP element, len: 112 aa;
                identical to CDS in other RLEP elements e.g. TR:O32969
                (MLCB22.36C)"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein MLCB1450.04"
                /protein_id="CAA22688.1"
                /db_xref="GI:4154039"
                /translation="MTTPTPQGHDMHTKTPLPRGANNYPHTHACIDIAFSTAQVPSPW
                HHQHVDQAAPTTDMLTCAALIVSTAAKHTKPHRKQAVSHPPTKTPQHSKTRQQDSGLP
                PTTTEHIGQS"
gene            complement(3504. .5129)
                /gene="groEL-2"
CDS             complement(3504. .5129)
                /gene="groEL-2"
                /note="MLCB1450.05c, groEL-2, 60 kD chaperonin, len: 541
                aa; identical to CH62_MYCLE, and highly similar to many
                e.g. CH60_BACSU 60 KD chaperonin (543 aa), fasta scores;
                opt: 2217 z-score: 2520.3 E(): 0, 63.3% identity in 539 aa
                overlap. Equivalent to M.tuberculosis CH62_MYCTU Rv0440
                (94.8% identity in 541 aa overlap). Contains PS00296
                Chaperonins cpn60 signature and Pfam match to entry
                PF00118 cpn60_TCP1, TCP-1 /cpn60 chaperonin family"
                /codon_start=1
                /transl_table=11
                /product="60 kD chaperonin"
                /protein_id="CAA22689.1"
                /db_xref="GI:4154040"
                /translation="MAKTIAYDEEARRGLERGLNSLADAVKVTLGPKGRNVVLEKKWG
                APTITNDGVSIAKEIELEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVKEG
                LRNVAAGANPLGLKRGIEKAVDKVTETLLKDAKEVETKEQIAATAAISAGDQSIGDLI
                AEAMDKVGNEGVITVEESNTFGLQLELTEGMRFDKGYISGYFVTDAERQEAVLEEPYI
                LLVSSKVSTVKDLLPLLEKVIQAGKSLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKA
                PGFGDRRKAMLQDMAILTGAQVISEEVGLTLENTDLSLLGKARKVVMTKDETTIVEGA
                GDTDAIAGRVAQIRTEIENSDSDYDREKLQERLAKLAGGVAVIKAGAATEVELKERKH
                RIEDAVRNAKAAVEEGIVAGGGVTLLQAAPALDKLKLTGDEATGANIVKVALEAPLKQ
                IAFNSGMEPGVVAEKVRNLSVGHGLNAATGEYEDLLKAGVADPVKVTRSALQNAASIA
                GLFLTTEAVVADKPEKTAAPASDPTGGMGGMDF"
misc_feature    complement(3564. .5066)
                /gene="groEL-2"
                /note="Pfam match to entry PF00118 cpn60_TCP1, TCP-1
                /cpn60 chaperonin family, score 778.80, E-value 2.2e-230"
misc_feature    complement(3888. .3923)
                /gene="groEL-2"
                /note="PS00296 Chaperonins cpn60 signature"
gene            complement(5338. .5907)
                /gene="MLCB1450.06c"
CDS             complement(5338. .5907)
                /gene="MLCB1450.06c"
                /note="MLCB1450.06c, probable transcriptional regulator,
                len: 189 aa; similar to hypothetical proteins from M.
                tuberculosis and S. coelicolor, and equivalent to
                M.tuberculosis putative transcriptional regulator Rv0067c
                (68.8% identity in 189 aa overlap). Contains probable
                helix-turn-helix motif at aa 35-56 (Score 1794, +5.30 SD)
                and Pfam match to entry PF00440 tetR, Bacterial regulatory
                proteins, tetR family"
                /codon_start=1
                /transl_table=11
                /product="putative transcriptional regulator"
                /protein_id="CAA22690.1"
                /db_xref="GI:4154041"
                /db_xref="SPTREMBL:Q9ZBM6"
                /translation="MAQPARPLRADAARNRARILGVAYEAFATEGLSVPIDEIARRAG
                VGAGTVYRHFPTKEALCAAVIGDRMPHLVDDGYVLLKSAGPGEALFTYLRSLVLHWGA
                TDRGLVDALAGAGGIDILGVAPDAEDAFLTILSDLLYAAQYAGTARTDVGVREVKSIL
                VGCQAMEAYNSALAERVTDVVVDGLRAAR"

  Query Match             64.0%;  Score 16;  DB 3;  Length 38065;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db  14193 CAGGTGCAGGTCAGCA 14208


RESULT  30
AC091120
LOCUS       AC091120    39520 bp    DNA             HTG     30-MAR-2001
DEFINITION  Leishmania major chromosome 2 clone L6541 strain Friedlin, ***
            SEQUENCING IN PROGRESS ***, 3 ordered pieces.
ACCESSION   AC091120
VERSION     AC091120.1  GI:13489160
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 39520)
  AUTHORS   Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K.,
            Robertson,L., Vogt,C., Marty,A., Cawthra,J., Seyler,A., Munden,H.,
            Nelson,S., Mack,J. and Pentony,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    27709: contig of 27709 bp in length
            *                   gap of unknown length
```

```
            *    27710    38388: contig of 10679 bp in length
            *                gap of unknown length
            *    38389    39520: contig of 1132 bp in length.
FEATURES             Location/Qualifiers
     source          1. .39520
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="2"
                     /clone="L6541"
BASE COUNT      6965 a  12678 c  12387 g   7487 t      3 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 78;  Length 39520;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db    1027 CAGGTGCAGGTCAGCA 1042


RESULT  31
AC004508
LOCUS       AC004508    71230 bp    DNA             PRI       30-MAR-1998
DEFINITION  Homo sapiens chromosome 21, P1 clone LBL#3 (LBNL H09), complete
            sequence.
ACCESSION   AC004508 L35684 L35685 L35686 L35687 AC000939 AC000938 AC000940
            AC000937 L81573 L35689 L35690 L35691 L35692 L35693 L35694 L35695
            L35696 L35697 L35698 L35699 L42077 L35700 L35701
VERSION     AC004508.1  GI:2996630
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71230)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71230)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 71230)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .71230
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="LBL#3"
                     /note="LBNL H09"
     repeat_region   complement(472. .788)
                     /rpt_family="MLT1a"
     repeat_region   1408. .1743
                     /rpt_family="Alu"
     misc_feature    4057. .4515
                     /note="99% identity zh62d01.r1"
                     /db_xref="dbEST:W86547"
     misc_feature    complement(4506. .4622)
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(4646. .5114)
                     /note="100% identity zh62d01.s1"
                     /db_xref="dbEST:W86451"
     STS             complement(4646. .5114)
                     /standard_name="SHGC-52452"
                     /note="100% identity"
                     /db_xref="dbSTS:G33232"
     repeat_region   complement(5148. .5285)
                     /rpt_family="Alu"
     repeat_region   complement(5788. .5902)
                     /rpt_family="MER44C"
     repeat_region   complement(5974. .6182)
                     /rpt_family="MER44C"
     repeat_region   complement(6153. .6473)
                     /rpt_family="Alu"
     repeat_region   complement(6485. .6565)
                     /rpt_family="MER44C"
     repeat_region   complement(6779. .6989)
                     /rpt_family="MER46"
     repeat_region   complement(7235. .8312)
                     /rpt_family="L1"
     repeat_region   complement(8410. .8569)
                     /rpt_family="L1"
     repeat_region   8577. .8987
                     /rpt_family="THE1"
     repeat_region   complement(9020. .9511)
                     /rpt_family="L1"
     repeat_region   9872. .10296
                     /rpt_family="THE1"
     repeat_region   10297. .11909
                     /rpt_family="MSTAR"
     repeat_region   11929. .12309
                     /rpt_family="THE1"
     repeat_region   12868. .13131
                     /rpt_family="Alu"
     repeat_region   complement(13267. .13635)
                     /rpt_family="L1"
     prim_transcript join(14992. .15122,21091. .21151,22200. .22263,25113. .25251,
                     33644. .33721)
                     /standard_name="yj99d04"
                     /note="98% identity R74138"
     misc_feature    14993. .15116
                     /note="GRAIL 2 excellent exon, frame 0"
     prim_transcript join(15038. .15122,21091. .21151,22209. .22263,25113. .25251,
                     33644. .33885)
                     /standard_name="zu55b03"
                     /note="98% identity AA402132"
     repeat_region   complement(15991. .16164)
                     /rpt_family="MER42"
     repeat_region   complement(16400. .16743)
                     /rpt_family="MLT2B2"
     repeat_region   complement(16761. .17073)
                     /rpt_family="Alu"
     repeat_region   16768. .16787
                     /note="(T)20"
                     /rpt_type=tandem
                     /rpt_unit=T
     misc_feature    19581. .19626
                     /note="GRAIL 2 excellent exon, frame 0"
     repeat_region   20435. .20765
                     /rpt_family="Alu"
     repeat_region   23936. .24290
                     /rpt_family="Alu"
     repeat_region   24273. .24292
                     /note="(A)20"
                     /rpt_type=tandem
                     /rpt_unit=A
     misc_feature    25115. .25247
                     /note="GRAIL 2 excellent exon, frame 2"
     repeat_region   26872. .26965
                     /rpt_family="MIR"
     repeat_region   29019. .29360
```

```
                /rpt_family="Alu"
repeat_region   29356. .31388
                /rpt_family="L1"
repeat_region   31440. .32072
                /rpt_family="L1"
repeat_region   complement(32120. .32230)
                /rpt_family="L1"
repeat_region   32581. .32624
                /note="(AC)22"
                /rpt_type=tandem
                /rpt_unit=AC
repeat_region   complement(32668. .32776)
                /rpt_family="MER3"
misc_feature    32932. .33120
                /note="86% identity zh52f10.r1"
                /db_xref="dbEST:W78940"
repeat_region   complement(33133. .33199)
                /rpt_family="MER3"
repeat_region   complement(33644. .33883)
                /note="99% identity nf66d01.s1"
                /db_xref="dbEST:AA515600"
misc_feature    33648. .33737
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   35260. .35660
                /rpt_family="MLT1"
repeat_region   36246. .36533
                /rpt_family="Alu"
repeat_region   complement(39299. .39589)
                /rpt_family="Alu"
repeat_region   complement(39781. .39931)
                /rpt_family="Alu"
repeat_region   complement(41134. .41588)
                /rpt_family="MLT2"
repeat_region   41621. .41706
                /rpt_family="MLT1"
repeat_region   41748. .41771
                /note="(GTTT)6"
                /rpt_type=tandem
                /rpt_unit=GTTT
repeat_region   complement(41757. .42397)
                /rpt_family="L1"
repeat_region   42935. .43106
                /rpt_family="Alu"
misc_feature    complement(44760. .44802)
                /note="GRAIL 2 excellent exon, frame 2"
repeat_region   complement(49449. .49589)
                /rpt_family="MIR2"
repeat_region   complement(51869. .52337)
                /rpt_family="MLT1"
repeat_region   52773. .53082
                /rpt_family="Alu"
gene            <53503. .>67215
                /gene="MX2"
                /note="interferon-regulated resistance GTP-binding protein
                MXB (P78-related protein)"
                /evidence=experimental
CDS             join(<53503. .53668,54302. .54494,56530. .56664,58923. .59077,
                67077. .>67215)
                /gene="MX2"
                /note="Interferon-regulated resistance GTP-binding protein
                MXB (P78-related protein)"
                /codon_start=2
                /evidence=experimental
                /product="MX2"
                /protein_id="AAC08448.1"
                /db_xref="GI:2996631"
                /translation="FQQQPPPFGTVPPQMMFPPNWQGAEKDAAFLAKDFNFLTLNNQP
                PPGNRSQPRAMGPENNLYSQYEQKVRPCIDLIDSLRALGVEQDLALPAIAVIGDQSSG
                KSSVLEALSGVALPRGSGIVTRCPLVLKLKKQPCEAWAGRISYRNTELELQDPGQVEK
                EIHKAQNVMAGNGRGISHELISLEITSPEVPDLTIIDLPGITRVAVDNQPRDIGLQIK
                ALIKKYIQRQQTINLVVVPCNVDIATTEALSMAHEVDPEGDRTI"
repeat_region   57922. .58215
                /rpt_family="Alu"
repeat_region   59599. .59877
                /rpt_family="Alu"
repeat_region   60428. .60810
                /rpt_family="THE1"
repeat_region   61246. .61366
                /rpt_family="Alu"
```

```
  Query Match             64.0%;  Score 16;  DB 85;  Length 71230;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db  40455 GTCAGGTGCAGGTCAG 40470


RESULT  32
AC002088/c
LOCUS       AC002088    85633 bp    DNA             PRI       04-FEB-2000
DEFINITION  Human BAC clone CTB-13P7 from 7p15-p21, complete sequence.
ACCESSION   AC002088
VERSION     AC002088.1  GI:2085779
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 85633)
  AUTHORS   Waterston,B. and Scheet,P.
  TITLE     The sequence of H. sapiens BAC clone CTB-13P7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1997)
REFERENCE   3  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG013P07
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send an E-mail to
            egreen@nhgri.nih.gov
```

SOURCE INFORMATION:
This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of CTB-13P7; actual end is at 85633 of CTB-13P7. The orientation of this clone is unknown.

This clone contains sWSS2365 (NID:g1254783).

FEATURES Location/Qualifiers

source	1. .85633
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="CTB-13P7"
	/clone_lib="CITB-978SK-B"
	/map="7p15-p21"
repeat_region	complement(3420. .3446)
	/rpt_family="L1"
repeat_region	complement(4016. .4306)
	/rpt_family="ALU"
repeat_region	complement(6902. .6942)
	/rpt_family="L1"
repeat_region	13108. .13401
	/rpt_family="ALU"
repeat_region	complement(16600. .16982)
	/rpt_family="L1"
repeat_region	complement(17300. .17375)
	/rpt_family="ALU"
repeat_region	complement(17475. .17502)
	/rpt_family="L1"
repeat_region	17503. .17584
	/rpt_family="ALU"
repeat_region	17625. .17838
	/rpt_family="ALU"
repeat_region	17979. .18078
	/rpt_family="L1"
repeat_region	complement(18407. .18454)
	/rpt_family="L1"
repeat_region	complement(24978. .25240)
	/rpt_family="ALU"
repeat_region	25293. .25377
	/rpt_family="L1"
repeat_region	complement(25503. .25778)
	/rpt_family="L1"
repeat_region	complement(25779. .26070)
	/rpt_family="ALU"
repeat_region	complement(26076. .26884)
	/rpt_family="L1"
repeat_region	complement(26887. .27177)
	/rpt_family="ALU"
repeat_region	complement(27242. .27327)
	/rpt_family="L1"
repeat_region	complement(27436. .27612)
	/rpt_family="L1"
repeat_region	complement(28177. .28229)
	/rpt_family="L1"
repeat_region	complement(28230. .28517)
	/rpt_family="ALU"
repeat_region	complement(28518. .28570)
	/rpt_family="L1"
repeat_region	31781. .32083
	/rpt_family="MER"
repeat_region	32102. .32130
	/rpt_family="MER"
repeat_region	complement(34983. .35002)
	/rpt_family="L1"
repeat_region	complement(35680. .35721)
	/rpt_family="L1"
repeat_region	complement(38336. .38360)
	/rpt_family="L1"
repeat_region	complement(39172. .39464)
	/rpt_family="ALU"
repeat_region	complement(39570. .39698)
	/rpt_family="ALU"
repeat_region	complement(39954. .40230)
	/rpt_family="L1"
repeat_region	40007. .40195
	/rpt_family="L1"
repeat_region	complement(40341. .40901)
	/rpt_family="L1"
repeat_region	40345. .40893
	/rpt_family="L1"
repeat_region	complement(41771. .47939)
	/rpt_family="L1"
repeat_region	complement(47950. .48026)
	/rpt_family="MER"
repeat_region	complement(48181. .48428)
	/rpt_family="MER"
repeat_region	48233. .48439
	/rpt_family="MER"
repeat_region	complement(48619. .48655)
	/rpt_family="MER"
repeat_region	complement(48694. .48890)
	/rpt_family="MER"
repeat_region	48819. .48974
	/rpt_family="MER"
repeat_region	49030. .49292
	/rpt_family="MER"
repeat_region	49440. .49757
	/rpt_family="MER"
repeat_region	complement(49570. .49873)
	/rpt_family="MER"
repeat_region	51098. .51155
	/rpt_family="ALU"
repeat_region	51181. .51230
	/rpt_family="ALU"
repeat_region	complement(51563. .51830)
	/rpt_family="ALU"
repeat_region	complement(52400. .52441)
	/rpt_family="L1"
repeat_region	complement(53064. .53085)
	/rpt_family="L1"
repeat_region	complement(53087. .53384)
	/rpt_family="ALU"
repeat_region	complement(53410. .53435)
	/rpt_family="L1"
repeat_region	complement(53436. .53545)
	/rpt_family="ALU"
repeat_region	complement(53554. .53709)
	/rpt_family="ALU"
repeat_region	complement(60972. .60992)
	/rpt_family="L1"
repeat_region	complement(64802. .64824)
	/rpt_family="L1"
repeat_region	complement(66753. .67258)
	/rpt_family="MER"
repeat_region	complement(72167. .72185)
	/rpt_family="L1"
repeat_region	72879. .73153
	/rpt_family="ALU"
repeat_region	complement(82252. .82468)
	/rpt_family="MER"
repeat_region	82281. .82473
	/rpt_family="MER"
repeat_region	82600. .82926
	/rpt_family="MER"
repeat_region	complement(83480. .83780)

```
  Query Match             64.0%;  Score 16;  DB 85;  Length 85633;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db  22093 CAGGTGCAGGTCAGCA 22078


RESULT  33
AC007555/c
LOCUS       AC007555   109973 bp    DNA             HTG       24-AUG-2000
DEFINITION  Homo sapiens chromosome 7 clone RP5-925B6, WORKING DRAFT SEQUENCE,
            2 unordered pieces.
ACCESSION   AC007555
VERSION     AC007555.2  GI:9887766
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109973)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109973)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 24, 2000 this sequence version replaced gi:4827326.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_DJ0925B06
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 79%
            Sequencing vector: plasmid; 21%
            Chemistry: Dye-primer ET; 79% of reads
            Chemistry: Dye-terminator Big Dye; 21% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 109177 bases at least Q40
            Consensus quality: 109655 bases at least Q30
            Consensus quality: 109773 bases at least Q20
            Insert size: 201000; agarose-fp
            Insert size: 0; sum-of-contigs
            Quality coverage:  in Q20 bases; agarose-fp
            Quality coverage:  in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25831: contig of 25831 bp in length
            *    25832    25931: gap of unknown length
            *    25932   109973: contig of 84042 bp in length.
FEATURES             Location/Qualifiers
     source          1. .109973
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP5-925B6"
     misc_feature    1. .25831
                     /note="assembly_name:Contig5
                     clone_end:SP6
                     vector_side:left"
     misc_feature    25932. .109973
                     /note="assembly_name:Contig6
                     clone_end:T7
                     vector_side:left"
BASE COUNT    34650 a  19714 c  19631 g  35877 t    101 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 60;  Length 109973;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db  91300 CAGGTGCAGGTCAGCA 91285


RESULT  34
HSMX1_4
WPCOMMENT
  Sequence split into 7 fragments  LOCUS HSMX1 Accession AJ011929
     Fragment Name      Begin     End
     HSMX1_0                1    110000
     HSMX1_1           100001    210000
     HSMX1_2           200001    310000
     HSMX1_3           300001    410000
     HSMX1_4           400001    510000
     HSMX1_5           500001    610000
     HSMX1_6           600001    613769
Continuation (5 of 7) of HSMX1 from base 400001 (AJ011929 Homo sapiens chromosome 21
s. 1/2000)


  Query Match             64.0%;  Score 16;  DB 84;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db  84114 GTCAGGTGCAGGTCAG 84129


RESULT  35
AC073052
LOCUS       AC073052   147129 bp    DNA             HTG       13-AUG-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-92F20, WORKING DRAFT SEQUENCE,
            27 unordered pieces.
ACCESSION   AC073052
VERSION     AC073052.2  GI:9799813
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147129)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147129)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 13, 2000 this sequence version replaced gi:8348014.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0092F20
```

```
-------------- Summary Statistics  --------------
Sequencing vector: M13; 93%
Sequencing vector: plasmid; 7%
Chemistry: Dye-primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 134469 bases at least Q40
Consensus quality: 137723 bases at least Q30
Consensus quality: 139595 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 144529; sum-of-contigs
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.95 in Q20 bases; sum-of-contigs
------------------------------------------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     1571: contig of 1571 bp in length
*      1572     1671: gap of unknown length
*      1672     2958: contig of 1287 bp in length
*      2959     3058: gap of unknown length
*      3059     4454: contig of 1396 bp in length
*      4455     4554: gap of unknown length
*      4555     5616: contig of 1062 bp in length
*      5617     5716: gap of unknown length
*      5717     7860: contig of 2144 bp in length
*      7861     7960: gap of unknown length
*      7961    11316: contig of 3356 bp in length
*     11317    11416: gap of unknown length
*     11417    14833: contig of 3417 bp in length
*     14834    14933: gap of unknown length
*     14934    18348: contig of 3415 bp in length
*     18349    18448: gap of unknown length
*     18449    22186: contig of 3738 bp in length
*     22187    22286: gap of unknown length
*     22287    25357: contig of 3071 bp in length
*     25358    25457: gap of unknown length
*     25458    28970: contig of 3513 bp in length
*     28971    29070: gap of unknown length
*     29071    31930: contig of 2860 bp in length
*     31931    32030: gap of unknown length
*     32031    35417: contig of 3387 bp in length
*     35418    35517: gap of unknown length
*     35518    39647: contig of 4130 bp in length
*     39648    39747: gap of unknown length
*     39748    43858: contig of 4111 bp in length
*     43859    43958: gap of unknown length
*     43959    49618: contig of 5660 bp in length
*     49619    49718: gap of unknown length
*     49719    54322: contig of 4604 bp in length
*     54323    54422: gap of unknown length
*     54423    61625: contig of 7203 bp in length
*     61626    61725: gap of unknown length
*     61726    65884: contig of 4159 bp in length
*     65885    65984: gap of unknown length
*     65985    71895: contig of 5911 bp in length
*     71896    71995: gap of unknown length
*     71996    78239: contig of 6244 bp in length
*     78240    78339: gap of unknown length
*     78340    83879: contig of 5540 bp in length
*     83880    83979: gap of unknown length
*     83980    90865: contig of 6886 bp in length
*     90866    90965: gap of unknown length
*     90966    98738: contig of 7773 bp in length
*     98739    98838: gap of unknown length
*     98839   110905: contig of 12067 bp in length
*    110906   111005: gap of unknown length
*    111006   124936: contig of 13931 bp in length
*    124937   125036: gap of unknown length
*    125037   147129: contig of 22093 bp in length.
FEATURES             Location/Qualifiers
     source          1. .147129
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-92F20"
     misc_feature    1. .1571
                     /note="assembly_name:Contig14"
     misc_feature    1672. .2958
                     /note="assembly_name:Contig15"
     misc_feature    3059. .4454
                     /note="assembly_name:Contig16"
     misc_feature    4555. .5616
                     /note="assembly_name:Contig17"
     misc_feature    5717. .7860
                     /note="assembly_name:Contig19"
     misc_feature    7961. .11316
                     /note="assembly_name:Contig20"
     misc_feature    11417. .14833
                     /note="assembly_name:Contig21"
     misc_feature    14934. .18348
                     /note="assembly_name:Contig22"
     misc_feature    18449. .22186
                     /note="assembly_name:Contig23"
     misc_feature    22287. .25357
                     /note="assembly_name:Contig24"
     misc_feature    25458. .28970
                     /note="assembly_name:Contig25"
     misc_feature    29071. .31930
                     /note="assembly_name:Contig26"
     misc_feature    32031. .35417
                     /note="assembly_name:Contig27"
     misc_feature    35518. .39647
                     /note="assembly_name:Contig28"
     misc_feature    39748. .43858
                     /note="assembly_name:Contig29"
     misc_feature    43959. .49618
                     /note="assembly_name:Contig30"
     misc_feature    49719. .54322
                     /note="assembly_name:Contig31"
     misc_feature    54423. .61625
                     /note="assembly_name:Contig32"
     misc_feature    61726. .65884
                     /note="assembly_name:Contig33"
     misc_feature    65985. .71895
                     /note="assembly_name:Contig34"
     misc_feature    71996. .78239
                     /note="assembly_name:Contig35"
     misc_feature    78340. .83879
                     /note="assembly_name:Contig36
                     clone_end:T7
                     vector_side:right"
     misc_feature    83980. .90865
                     /note="assembly_name:Contig37"
     misc_feature    90966. .98738
                     /note="assembly_name:Contig38"
     misc_feature    98839. .110905
                     /note="assembly_name:Contig39"
     misc_feature    111006. .124936
                     /note="assembly_name:Contig40"
     misc_feature    125037. .147129
                     /note="assembly_name:Contig41"
BASE COUNT    44181 a  28681 c  26793 g  44841 t   2633 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 74;  Length 147129;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
```

```
          ||||||||||||||||
Db    6339 CAGGTGCAGGTCAGCA 6354


RESULT  36
AC027810/c
LOCUS       AC027810   155628 bp    DNA             HTG       26-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-357J12 map 18, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC027810
VERSION     AC027810.3  GI:8082020
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155628)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-357J12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155628)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7684514.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9097
                Center clone name: 357_J_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 146315 bases at least Q40
                Consensus quality: 151241 bases at least Q30
                Consensus quality: 153082 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 154328; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1068: contig of 1068 bp in length
            *      1069 1168: gap of      100 bp
            *      1169     2717: contig of 1549 bp in length
            *      2718 2817: gap of      100 bp
            *      2818     6040: contig of 3223 bp in length
            *      6041 6140: gap of      100 bp
            *      6141    10201: contig of 4061 bp in length
            *     10202 10301: gap of      100 bp
            *     10302    14615: contig of 4314 bp in length
            *     14616 14715: gap of      100 bp
            *     14716    21086: contig of 6371 bp in length
            *     21087 21186: gap of      100 bp
            *     21187    25767: contig of 4581 bp in length
            *     25768 25867: gap of      100 bp
            *     25868    32500: contig of 6633 bp in length
            *     32501 32600: gap of      100 bp
            *     32601    38651: contig of 6051 bp in length
            *     38652 38751: gap of      100 bp
            *     38752    50922: contig of 12171 bp in length
            *     50923 51022: gap of      100 bp
            *     51023    71014: contig of 19992 bp in length
            *     71015 71114: gap of      100 bp
            *     71115    93563: contig of 22449 bp in length
            *     93564 93663: gap of      100 bp
            *     93664   123080: contig of 29417 bp in length
            *    123081 123180: gap of      100 bp
            *    123181   155628: contig of 32448 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155628
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-357J12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1068
                     /note="assembly_fragment"
     misc_feature    1169. .2717
                     /note="assembly_fragment"
     misc_feature    2818. .6040
                     /note="assembly_fragment"
     misc_feature    6141. .10201
                     /note="assembly_fragment"
     misc_feature    10302. .14615
                     /note="assembly_fragment"
     misc_feature    14716. .21086
                     /note="assembly_fragment"
     misc_feature    21187. .25767
                     /note="assembly_fragment"
     misc_feature    25868. .32500
                     /note="assembly_fragment"
     misc_feature    32601. .38651
                     /note="assembly_fragment"
     misc_feature    38752. .50922
                     /note="assembly_fragment"
     misc_feature    51023. .71014
                     /note="assembly_fragment"
     misc_feature    71115. .93563
                     /note="assembly_fragment"
     misc_feature    93664. .123080
                     /note="assembly_fragment"
     misc_feature    123181. .155628
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    37665 a  38986 c  40608 g  37067 t   1302 others
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 71;  Length 155628;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db   95800 GTCAGGTGCAGGTCAG 95785


RESULT  37
AC073247
LOCUS       AC073247   156119 bp    DNA             HTG       12-AUG-2000
DEFINITION  Homo sapiens chromosome 15 clone CTD-2006N8, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC073247
VERSION     AC073247.2  GI:8571764
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156119)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156119)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 18, 2000 this sequence version replaced gi:8468936.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_MS2006N08
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 139572 bases at least Q40
            Consensus quality: 144812 bases at least Q30
            Consensus quality: 147446 bases at least Q20
            Insert size: 160000; agarose-fp
            Insert size: 154019; sum-of-contigs
            Quality coverage: 4.17 in Q20 bases; agarose-fp
            Quality coverage: 4.44 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1286: contig of 1286 bp in length
            *     1287     1386: gap of unknown length
            *     1387     3603: contig of 2217 bp in length
            *     3604     3703: gap of unknown length
            *     3704     6043: contig of 2340 bp in length
            *     6044     6143: gap of unknown length
            *     6144     9354: contig of 3211 bp in length
            *     9355     9454: gap of unknown length
            *     9455    12077: contig of 2623 bp in length
            *    12078    12177: gap of unknown length
            *    12178    15985: contig of 3808 bp in length
            *    15986    16085: gap of unknown length
            *    16086    19801: contig of 3716 bp in length
            *    19802    19901: gap of unknown length
            *    19902    25990: contig of 6089 bp in length
            *    25991    26090: gap of unknown length
            *    26091    30598: contig of 4508 bp in length
            *    30599    30698: gap of unknown length
            *    30699    36263: contig of 5565 bp in length
            *    36264    36363: gap of unknown length
            *    36364    42783: contig of 6420 bp in length
            *    42784    42883: gap of unknown length
            *    42884    49478: contig of 6595 bp in length
            *    49479    49578: gap of unknown length
            *    49579    56751: contig of 7173 bp in length
            *    56752    56851: gap of unknown length
            *    56852    61885: contig of 5034 bp in length
            *    61886    61985: gap of unknown length
            *    61986    69384: contig of 7399 bp in length
            *    69385    69484: gap of unknown length
            *    69485    76911: contig of 7427 bp in length
            *    76912    77011: gap of unknown length
            *    77012    86274: contig of 9263 bp in length
            *    86275    86374: gap of unknown length
            *    86375    96575: contig of 10201 bp in length
            *    96576    96675: gap of unknown length
            *    96676   109456: contig of 12781 bp in length
            *   109457   109556: gap of unknown length
            *   109557   123883: contig of 14327 bp in length
            *   123884   123983: gap of unknown length
            *   123984   138107: contig of 14124 bp in length
            *   138108   138207: gap of unknown length
            *   138208   156119: contig of 17912 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156119
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="CTD-2006N8"
BASE COUNT    42637 a  33200 c  33979 g  44179 t   2124 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 74;  Length 156119;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   90601 CAGGTGCAGGTCAGCA 90616


RESULT  38
AC040930
LOCUS       AC040930   162043 bp    DNA             HTG       03-FEB-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD-2006N8 map 15, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC040930
VERSION     AC040930.3  GI:12658160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162043)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-2006N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162043)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
```

```
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:7705184.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8890
                Center clone name: 2006_N_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 47% of reads
                Sequencing vector: Plasmid; n/a; 53% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158666 bases at least Q40
                Consensus quality: 159912 bases at least Q30
                Consensus quality: 160439 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161043; sum-of-contigs
                Quality coverage: 10.5 in Q20 bases; agarose-fp
                Quality coverage: 10.2 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      840: contig of 840 bp in length
            *      841 940: gap of      100 bp
            *      941     1652: contig of 712 bp in length
            *     1653 1752: gap of      100 bp
            *     1753     2845: contig of 1093 bp in length
            *     2846 2945: gap of      100 bp
            *     2946     4039: contig of 1094 bp in length
            *     4040 4139: gap of      100 bp
            *     4140     7025: contig of 2886 bp in length
            *     7026 7125: gap of      100 bp
            *     7126    10842: contig of 3717 bp in length
            *    10843 10942: gap of      100 bp
            *    10943    18373: contig of 7431 bp in length
            *    18374 18473: gap of      100 bp
            *    18474    28208: contig of 9735 bp in length
            *    28209 28308: gap of      100 bp
            *    28309    48596: contig of 20288 bp in length
            *    48597 48696: gap of      100 bp
            *    48697    76612: contig of 27916 bp in length
            *    76613 76712: gap of      100 bp
            *    76713   162043: contig of 85331 bp in length.
FEATURES            Location/Qualifiers
     source         1. .162043
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="15"
                    /map="15"
                    /clone="CTD-2006N8"
                    /clone_lib="CITD Human BAC"
     misc_feature   1. .840
                    /note="assembly_fragment"
     misc_feature   941. .1652
                    /note="assembly_fragment"
     misc_feature   1753. .2845
                    /note="assembly_fragment"
     misc_feature   2946. .4039
                    /note="assembly_fragment"
     misc_feature   4140. .7025
                    /note="assembly_fragment"
     misc_feature   7126. .10842
                    /note="assembly_fragment"
     misc_feature   10943. .18373
                    /note="assembly_fragment"
     misc_feature   18474. .28208
                    /note="assembly_fragment"
     misc_feature   28309. .48596
                    /note="assembly_fragment"
     misc_feature   48697. .76612
                    /note="assembly_fragment"
     misc_feature   76713. .162043
                    /note="assembly_fragment"
BASE COUNT    44171 a  35675 c  34546 g  46650 t   1001 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 71;  Length 162043;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   82976 CAGGTGCAGGTCAGCA 82991


RESULT  39
AC068538/c
LOCUS       AC068538   175459 bp    DNA             HTG       13-AUG-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-141G4, WORKING DRAFT SEQUENCE,
            29 unordered pieces.
ACCESSION   AC068538
VERSION     AC068538.5  GI:9795677
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175459)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175459)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 11, 2000 this sequence version replaced gi:9690368.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0141G04
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
```

```
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158616 bases at least Q40
Consensus quality: 162423 bases at least Q30
Consensus quality: 164318 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 172659; sum-of-contigs
Quality coverage: 4.51 in Q20 bases; agarose-fp
Quality coverage: 4.87 in Q20 bases; sum-of-contigs
--------------------------------------------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1355: contig of 1355 bp in length
*     1356     1455: gap of unknown length
*     1456     3082: contig of 1627 bp in length
*     3083     3182: gap of unknown length
*     3183     4263: contig of 1081 bp in length
*     4264     4363: gap of unknown length
*     4364     5626: contig of 1263 bp in length
*     5627     5726: gap of unknown length
*     5727     7533: contig of 1807 bp in length
*     7534     7633: gap of unknown length
*     7634     9202: contig of 1569 bp in length
*     9203     9302: gap of unknown length
*     9303    11061: contig of 1759 bp in length
*    11062    11161: gap of unknown length
*    11162    14017: contig of 2856 bp in length
*    14018    14117: gap of unknown length
*    14118    16928: contig of 2811 bp in length
*    16929    17028: gap of unknown length
*    17029    19922: contig of 2894 bp in length
*    19923    20022: gap of unknown length
*    20023    23986: contig of 3964 bp in length
*    23987    24086: gap of unknown length
*    24087    27864: contig of 3778 bp in length
*    27865    27964: gap of unknown length
*    27965    31695: contig of 3731 bp in length
*    31696    31795: gap of unknown length
*    31796    34671: contig of 2876 bp in length
*    34672    34771: gap of unknown length
*    34772    39459: contig of 4688 bp in length
*    39460    39559: gap of unknown length
*    39560    43565: contig of 4006 bp in length
*    43566    43665: gap of unknown length
*    43666    47318: contig of 3653 bp in length
*    47319    47418: gap of unknown length
*    47419    53986: contig of 6568 bp in length
*    53987    54086: gap of unknown length
*    54087    61186: contig of 7100 bp in length
*    61187    61286: gap of unknown length
*    61287    66331: contig of 5045 bp in length
*    66332    66431: gap of unknown length
*    66432    72447: contig of 6016 bp in length
*    72448    72547: gap of unknown length
*    72548    80701: contig of 8154 bp in length
*    80702    80801: gap of unknown length
*    80802    89408: contig of 8607 bp in length
*    89409    89508: gap of unknown length
*    89509   100140: contig of 10632 bp in length
*   100141   100240: gap of unknown length
*   100241   109860: contig of 9620 bp in length
*   109861   109960: gap of unknown length
*   109961   120505: contig of 10545 bp in length
*   120506   120605: gap of unknown length
*   120606   135065: contig of 14460 bp in length
*   135066   135165: gap of unknown length
*   135166   153777: contig of 18612 bp in length
*   153778   153877: gap of unknown length
*   153878   175459: contig of 21582 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175459
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-141G4"
     misc_feature    1. .1355
                     /note="assembly_name:Contig9"
     misc_feature    1456. .3082
                     /note="assembly_name:Contig11"
     misc_feature    3183. .4263
                     /note="assembly_name:Contig13"
     misc_feature    4364. .5626
                     /note="assembly_name:Contig14"
     misc_feature    5727. .7533
                     /note="assembly_name:Contig15"
     misc_feature    7634. .9202
                     /note="assembly_name:Contig16
                     clone_end:T7
                     vector_side:left"
     misc_feature    9303. .11061
                     /note="assembly_name:Contig17"
     misc_feature    11162. .14017
                     /note="assembly_name:Contig18"
     misc_feature    14118. .16928
                     /note="assembly_name:Contig19"
     misc_feature    17029. .19922
                     /note="assembly_name:Contig20"
     misc_feature    20023. .23986
                     /note="assembly_name:Contig21"
     misc_feature    24087. .27864
                     /note="assembly_name:Contig22"
     misc_feature    27965. .31695
                     /note="assembly_name:Contig23"
     misc_feature    31796. .34671
                     /note="assembly_name:Contig24"
     misc_feature    34772. .39459
                     /note="assembly_name:Contig25"
     misc_feature    39560. .43565
                     /note="assembly_name:Contig26"
     misc_feature    43666. .47318
                     /note="assembly_name:Contig27"
     misc_feature    47419. .53986
                     /note="assembly_name:Contig28"
     misc_feature    54087. .61186
                     /note="assembly_name:Contig29"
     misc_feature    61287. .66331
                     /note="assembly_name:Contig30"
     misc_feature    66432. .72447
                     /note="assembly_name:Contig31
                     clone_end:SP6
                     vector_side:right"
     misc_feature    72548. .80701
                     /note="assembly_name:Contig32"
     misc_feature    80802. .89408
                     /note="assembly_name:Contig33"
     misc_feature    89509. .100140
                     /note="assembly_name:Contig34"
     misc_feature    100241. .109860
                     /note="assembly_name:Contig35"
     misc_feature    109961. .120505
                     /note="assembly_name:Contig36"
     misc_feature    120606. .135065
                     /note="assembly_name:Contig37"
     misc_feature    135166. .153777
                     /note="assembly_name:Contig38"
     misc_feature    153878. .175459
                     /note="assembly_name:Contig39"
BASE COUNT    47828 a  38505 c  38363 g  47961 t   2802 others
```

ORIGIN

Query Match 64.0%; Score 16; DB 73; Length 175459;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   79767 CAGGTGCAGGTCAGCA 79752
```

RESULT 40
AC005412/c
LOCUS AC005412 183700 bp DNA PRI 22-NOV-2000
DEFINITION Homo sapiens chromosome 17, clone hRPK.22_N_12, complete sequence.
ACCESSION AC005412
VERSION AC005412.6 GI:11128436
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183700)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.22_N_12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183700)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 183700)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 183700)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 9, 2000 this sequence version replaced gi:5931398.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES Location/Qualifiers

source	1. .183700
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="17"
	/map="17"
	/clone="hRPK.22_N_12"
	/clone_lib="RPCI-11 Human Male BAC"
repeat_region	1. .70
	/rpt_family="MER41D"
repeat_region	784. .1084
	/rpt_family="AluY"
repeat_region	1320. .1637
	/rpt_family="AluJo"
repeat_region	complement(1664. .1959)
	/rpt_family="AluSx"
repeat_region	1961. .2074
	/rpt_family="L2"
repeat_region	2283. .2355
	/rpt_family="C-rich"
repeat_region	complement(2369. .2774)
	/rpt_family="L1M4"
repeat_region	2775. .3066
	/rpt_family="AluSg"
repeat_region	complement(3067. .3077)
	/rpt_family="L1M4"
repeat_region	3549. .3814
	/rpt_family="AluSq"
repeat_region	complement(3867. .3949)
	/rpt_family="AluJo"
repeat_region	complement(3950. .4247)
	/rpt_family="AluJb"
repeat_region	complement(4248. .4361)
	/rpt_family="AluJo"
repeat_region	complement(4362. .4490)
	/rpt_family="L1M4"
repeat_region	complement(4493. .4589)
	/rpt_family="L1M4"
repeat_region	4590. .4888
	/rpt_family="AluSx"
repeat_region	complement(4889. .5305)
	/rpt_family="L1M4"
repeat_region	complement(5328. .5623)
	/rpt_family="AluSx"
repeat_region	complement(5876. .6164)
	/rpt_family="AluSg"
repeat_region	6196. .6509
	/rpt_family="AluSg"
repeat_region	6525. .6580
	/rpt_family="AT_rich"
repeat_region	complement(6671. .6784)
	/rpt_family="L2"
repeat_region	complement(6986. .7137)
	/rpt_family="AluSg"
repeat_region	complement(7138. .7447)
	/rpt_family="AluSx"
repeat_region	complement(7448. .7582)
	/rpt_family="AluSg"
repeat_region	7808. .8106

```
                /rpt_family="AluJb"
     unsure     7825. .8084
                /note="Single-stranded terminator coverage."
     repeat_region 8121. .8476
                /rpt_family="L1MB5"
     repeat_region 8496. .8677
                /rpt_family="(TATATG)n"
     unsure     8503. .8791
                /note="Single-stranded terminator coverage."
     repeat_region 8689. .8870
                /rpt_family="(TA)n"
     unsure     8736
                /note="Probably T, possibly C."
     repeat_region 8877. .9144
                /rpt_family="(TATATG)n"
     repeat_region 9147. .9210
                /rpt_family="(TA)n"
     repeat_region 9216. .9395
                /rpt_family="(CATATA)n"
     unsure     complement(9240. .9299)
                /note="Single-stranded terminator coverage."
     repeat_region 9458. .9530
                /rpt_family="L1MB5"
     repeat_region 9536. .9848
                /rpt_family="AluSx"
     repeat_region 9878. .10115
                /rpt_family="L2"
     repeat_region 10128. .10357
                /rpt_family="MER20"
     repeat_region 10432. .10458
                /rpt_family="AT_rich"
     repeat_region complement(10511. .10567)
                /rpt_family="AluJ/FLAM"
     repeat_region 10612. .10908
                /rpt_family="AluJb"
     repeat_region 10913. .11210
                /rpt_family="AluSp"
     repeat_region 11216. .11512
                /rpt_family="AluSx"
     repeat_region 11529. .11553
                /rpt_family="(TGAA)n"
     repeat_region complement(11647. .11759)
                /rpt_family="L2"
     repeat_region complement(11816. .12159)
                /rpt_family="L2"
     repeat_region complement(13008. .13306)
                /rpt_family="AluSp"
     repeat_region 13355. .13490
                /rpt_family="AluSx"
     repeat_region 13491. .13789
                /rpt_family="AluSc"
     repeat_region 13790. .13956
                /rpt_family="AluSx"
     repeat_region complement(14078. .14206)
                /rpt_family="MIR"
     repeat_region complement(14220. .14276)
                /rpt_family="L2"
     repeat_region complement(14277. .14309)
                /rpt_family="MIR"
     repeat_region complement(14310. .14594)
                /rpt_family="AluSp"
     repeat_region complement(14604. .14912)
                /rpt_family="AluJo"
     repeat_region complement(14913. .15045)
                /rpt_family="MIR"
     repeat_region complement(15061. .15157)
                /rpt_family="MIR"
     repeat_region complement(15517. .15637)

Query Match           64.0%;  Score 16;  DB 86;  Length 183700;
Best Local Similarity 100.0%;  Pred. No. 46;
Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db 177274 GTCAGGTGCAGGTCAG 177259


RESULT  41
AC025339/c
LOCUS       AC025339   185492 bp    DNA             HTG       08-MAR-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-676D22 map 2, WORKING DRAFT
            SEQUENCE, 51 unordered pieces.
ACCESSION   AC025339
VERSION     AC025339.1  GI:7210047
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185492)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-676D22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185492)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6951
                Center clone name: 676_D_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 144276 bases at least Q40
                Consensus quality: 161970 bases at least Q30
                Consensus quality: 171080 bases at least Q20
                Insert size: 194000; agarose-fp
                Insert size: 180492; sum-of-contigs
                Quality coverage: 2.9 in Q20 bases; agarose-fp
                Quality coverage: 3.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 51 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1026: contig of 1026 bp in length
*     1027 1126: gap of      100 bp
*     1127     2310: contig of 1184 bp in length
*     2311 2410: gap of      100 bp
*     2411     3533: contig of 1123 bp in length
*     3534 3633: gap of      100 bp
*     3634     4974: contig of 1341 bp in length
*     4975 5074: gap of      100 bp
*     5075     6187: contig of 1113 bp in length
*     6188 6287: gap of      100 bp
*     6288     7383: contig of 1096 bp in length
*     7384 7483: gap of      100 bp
*     7484     8508: contig of 1025 bp in length
*     8509 8608: gap of      100 bp
*     8609     9799: contig of 1191 bp in length
*     9800 9899: gap of      100 bp
*     9900    11275: contig of 1376 bp in length
*    11276 11375: gap of      100 bp
*    11376    12804: contig of 1429 bp in length
*    12805 12904: gap of      100 bp
*    12905    14246: contig of 1342 bp in length
*    14247 14346: gap of      100 bp
*    14347    15544: contig of 1198 bp in length
*    15545 15644: gap of      100 bp
*    15645    17181: contig of 1537 bp in length
*    17182 17281: gap of      100 bp
*    17282    19392: contig of 2111 bp in length
*    19393 19492: gap of      100 bp
*    19493    21710: contig of 2218 bp in length
*    21711 21810: gap of      100 bp
*    21811    24149: contig of 2339 bp in length
*    24150 24249: gap of      100 bp
*    24250    26685: contig of 2436 bp in length
*    26686 26785: gap of      100 bp
*    26786    28327: contig of 1542 bp in length
*    28328 28427: gap of      100 bp
*    28428    31535: contig of 3108 bp in length
*    31536 31635: gap of      100 bp
*    31636    34165: contig of 2530 bp in length
*    34166 34265: gap of      100 bp
*    34266    37131: contig of 2866 bp in length
*    37132 37231: gap of      100 bp
*    37232    40119: contig of 2888 bp in length
*    40120 40219: gap of      100 bp
*    40220    43234: contig of 3015 bp in length
*    43235 43334: gap of      100 bp
*    43335    46125: contig of 2791 bp in length
*    46126 46225: gap of      100 bp
*    46226    48748: contig of 2523 bp in length
*    48749 48848: gap of      100 bp
*    48849    53250: contig of 4402 bp in length
*    53251 53350: gap of      100 bp
*    53351    57172: contig of 3822 bp in length
*    57173 57272: gap of      100 bp
*    57273    60281: contig of 3009 bp in length
*    60282 60381: gap of      100 bp
*    60382    63484: contig of 3103 bp in length
*    63485 63584: gap of      100 bp
*    63585    67025: contig of 3441 bp in length
*    67026 67125: gap of      100 bp
*    67126    70475: contig of 3350 bp in length
*    70476 70575: gap of      100 bp
*    70576    74638: contig of 4063 bp in length
*    74639 74738: gap of      100 bp
*    74739    78257: contig of 3519 bp in length
*    78258 78357: gap of      100 bp
*    78358    83359: contig of 5002 bp in length
*    83360 83459: gap of      100 bp
*    83460    86986: contig of 3527 bp in length
*    86987 87086: gap of      100 bp
*    87087    90187: contig of 3101 bp in length
*    90188 90287: gap of      100 bp
*    90288    94735: contig of 4448 bp in length
*    94736 94835: gap of      100 bp
*    94836    99768: contig of 4933 bp in length
*    99769 99868: gap of      100 bp
*    99869   104278: contig of 4410 bp in length
*   104279 104378: gap of      100 bp
*   104379   110808: contig of 6430 bp in length
*   110809 110908: gap of      100 bp
*   110909   116478: contig of 5570 bp in length
*   116479 116578: gap of      100 bp
*   116579   122443: contig of 5865 bp in length
*   122444 122543: gap of      100 bp
*   122544   126331: contig of 3788 bp in length
*   126332 126431: gap of      100 bp
*   126432   131374: contig of 4943 bp in length
*   131375 131474: gap of      100 bp
*   131475   138177: contig of 6703 bp in length
*   138178 138277: gap of      100 bp
*   138278   144292: contig of 6015 bp in length
*   144293 144392: gap of      100 bp
*   144393   150325: contig of 5933 bp in length
*   150326 150425: gap of      100 bp
*   150426   156168: contig of 5743 bp in length
*   156169 156268: gap of      100 bp
*   156269   163498: contig of 7230 bp in length
*   163499 163598: gap of      100 bp
*   163599   173145: contig of 9547 bp in length
*   173146 173245: gap of      100 bp
*   173246   185492: contig of 12247 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185492
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-676D22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1026
                     /note="assembly_fragment"
     misc_feature    1127. .2310
                     /note="assembly_fragment"
     misc_feature    2411. .3533
                     /note="assembly_fragment"
     misc_feature    3634. .4974
                     /note="assembly_fragment"
     misc_feature    5075. .6187
                     /note="assembly_fragment"
     misc_feature    6288. .7383
                     /note="assembly_fragment"
     misc_feature    7484. .8508
                     /note="assembly_fragment"
     misc_feature    8609. .9799
                     /note="assembly_fragment"
     misc_feature    9900. .11275
                     /note="assembly_fragment"
     misc_feature    11376. .12804
                     /note="assembly_fragment"
     misc_feature    12905. .14246

  Query Match             64.0%;  Score 16;  DB 69;  Length 185492;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   43512 CAGGTGCAGGTCAGCA 43497


RESULT  42
AC064835
```

```
LOCUS       AC064835   202828 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-673P17, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC064835
VERSION     AC064835.3  GI:7839924
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202828)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202828)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 16, 2000 this sequence version replaced gi:7712280.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0673P17
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 195107 bases at least Q40
            Consensus quality: 197538 bases at least Q30
            Consensus quality: 198979 bases at least Q20
            Insert size: 199000; agarose-fp
            Insert size: 201728; sum-of-contigs
            Quality coverage: 5.33 in Q20 bases; agarose-fp
            Quality coverage: 5.29 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1109: contig of 1109 bp in length
            *     1110     1209: gap of unknown length
            *     1210     5935: contig of 4726 bp in length
            *     5936     6035: gap of unknown length
            *     6036    13020: contig of 6985 bp in length
            *    13021    13120: gap of unknown length
            *    13121    21028: contig of 7908 bp in length
            *    21029    21128: gap of unknown length
            *    21129    27413: contig of 6285 bp in length
            *    27414    27513: gap of unknown length
            *    27514    34551: contig of 7038 bp in length
            *    34552    34651: gap of unknown length
            *    34652    53143: contig of 18492 bp in length
            *    53144    53243: gap of unknown length
            *    53244    77201: contig of 23958 bp in length
            *    77202    77301: gap of unknown length
            *    77302   106156: contig of 28855 bp in length
            *   106157   106256: gap of unknown length
            *   106257   131644: contig of 25388 bp in length
            *   131645   131744: gap of unknown length
            *   131745   158698: contig of 26954 bp in length
            *   158699   158798: gap of unknown length
            *   158799   202828: contig of 44030 bp in length.

FEATURES             Location/Qualifiers
     source          1. .202828
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-673P17"
     misc_feature    1. .1109
                     /note="assembly_name:Contig14"
     misc_feature    1210. .5935
                     /note="assembly_name:Contig18"
     misc_feature    6036. .13020
                     /note="assembly_name:Contig19"
     misc_feature    13121. .21028
                     /note="assembly_name:Contig20"
     misc_feature    21129. .27413
                     /note="assembly_name:Contig21
                     clone_end:T7
                     vector_side:right"
     misc_feature    27514. .34551
                     /note="assembly_name:Contig22"
     misc_feature    34652. .53143
                     /note="assembly_name:Contig23"
     misc_feature    53244. .77201
                     /note="assembly_name:Contig24"
     misc_feature    77302. .106156
                     /note="assembly_name:Contig25"
     misc_feature    106257. .131644
                     /note="assembly_name:Contig26
                     clone_end:SP6
                     vector_side:left"
     misc_feature    131745. .158698
                     /note="assembly_name:Contig27"
     misc_feature    158799. .202828
                     /note="assembly_name:Contig28"
BASE COUNT    53011 a  46987 c  47342 g  54379 t   1109 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 72;  Length 202828;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   46988 CAGGTGCAGGTCAGCA 47003


RESULT  43
AC016186/c
LOCUS       AC016186   245499 bp    DNA             HTG       12-MAR-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-8I15 map 18, *** SEQUENCING
            IN PROGRESS ***, 70 unordered pieces.
ACCESSION   AC016186
VERSION     AC016186.3  GI:7230177
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245499)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-8I15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245499)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
```

```
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6649364.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4258
                Center clone name: 8_I_15
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 70 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1174: contig of 1174 bp in length
            *    1175 1274: gap of      100 bp
            *    1275     2609: contig of 1335 bp in length
            *    2610 2709: gap of      100 bp
            *    2710     3931: contig of 1222 bp in length
            *    3932 4031: gap of      100 bp
            *    4032     5121: contig of 1090 bp in length
            *    5122 5221: gap of      100 bp
            *    5222     6287: contig of 1066 bp in length
            *    6288 6387: gap of      100 bp
            *    6388     7686: contig of 1299 bp in length
            *    7687 7786: gap of      100 bp
            *    7787     8797: contig of 1011 bp in length
            *    8798 8897: gap of      100 bp
            *    8898    10536: contig of 1639 bp in length
            *   10537 10636: gap of      100 bp
            *   10637    11858: contig of 1222 bp in length
            *   11859 11958: gap of      100 bp
            *   11959    13375: contig of 1417 bp in length
            *   13376 13475: gap of      100 bp
            *   13476    15102: contig of 1627 bp in length
            *   15103 15202: gap of      100 bp
            *   15203    16445: contig of 1243 bp in length
            *   16446 16545: gap of      100 bp
            *   16546    17637: contig of 1092 bp in length
            *   17638 17737: gap of      100 bp
            *   17738    19309: contig of 1572 bp in length
            *   19310 19409: gap of      100 bp
            *   19410    21532: contig of 2123 bp in length
            *   21533 21632: gap of      100 bp
            *   21633    23166: contig of 1534 bp in length
            *   23167 23266: gap of      100 bp
            *   23267    25262: contig of 1996 bp in length
            *   25263 25362: gap of      100 bp
            *   25363    27315: contig of 1953 bp in length
            *   27316 27415: gap of      100 bp
            *   27416    28722: contig of 1307 bp in length
            *   28723 28822: gap of      100 bp
            *   28823    30596: contig of 1774 bp in length
            *   30597 30696: gap of      100 bp
            *   30697    33203: contig of 2507 bp in length
            *   33204 33303: gap of      100 bp
            *   33304    35408: contig of 2105 bp in length
            *   35409 35508: gap of      100 bp
            *   35509    37441: contig of 1933 bp in length
            *   37442 37541: gap of      100 bp
            *   37542    40401: contig of 2860 bp in length
            *   40402 40501: gap of      100 bp
            *   40502    42052: contig of 1551 bp in length
            *   42053 42152: gap of      100 bp
            *   42153    44294: contig of 2142 bp in length
            *   44295 44394: gap of      100 bp
            *   44395    46328: contig of 1934 bp in length
            *   46329 46428: gap of      100 bp
            *   46429    48649: contig of 2221 bp in length
            *   48650 48749: gap of      100 bp
            *   48750    51037: contig of 2288 bp in length
            *   51038 51137: gap of      100 bp
            *   51138    53348: contig of 2211 bp in length
            *   53349 53448: gap of      100 bp
            *   53449    55337: contig of 1889 bp in length
            *   55338 55437: gap of      100 bp
            *   55438    57028: contig of 1591 bp in length
            *   57029 57128: gap of      100 bp
            *   57129    59830: contig of 2702 bp in length
            *   59831 59930: gap of      100 bp
            *   59931    62172: contig of 2242 bp in length
            *   62173 62272: gap of      100 bp
            *   62273    64940: contig of 2668 bp in length
            *   64941 65040: gap of      100 bp
            *   65041    67862: contig of 2822 bp in length
            *   67863 67962: gap of      100 bp
            *   67963    70236: contig of 2274 bp in length
            *   70237 70336: gap of      100 bp
            *   70337    72533: contig of 2197 bp in length
            *   72534 72633: gap of      100 bp
            *   72634    74796: contig of 2163 bp in length
            *   74797 74896: gap of      100 bp
            *   74897    77899: contig of 3003 bp in length
            *   77900 77999: gap of      100 bp
            *   78000    81075: contig of 3076 bp in length
            *   81076 81175: gap of      100 bp
            *   81176    84346: contig of 3171 bp in length
            *   84347 84446: gap of      100 bp
            *   84447    87026: contig of 2580 bp in length
            *   87027 87126: gap of      100 bp
            *   87127    90859: contig of 3733 bp in length
            *   90860 90959: gap of      100 bp
            *   90960    93599: contig of 2640 bp in length
            *   93600 93699: gap of      100 bp
            *   93700    96946: contig of 3247 bp in length
            *   96947 97046: gap of      100 bp
            *   97047   100076: contig of 3030 bp in length
            *  100077 100176: gap of      100 bp
            *  100177   105333: contig of 5157 bp in length
            *  105334 105433: gap of      100 bp
            *  105434   109169: contig of 3736 bp in length
            *  109170 109269: gap of      100 bp
            *  109270   113940: contig of 4671 bp in length
            *  113941 114040: gap of      100 bp
            *  114041   119295: contig of 5255 bp in length
            *  119296 119395: gap of      100 bp
            *  119396   120878: contig of 1483 bp in length
            *  120879 120978: gap of      100 bp
            *  120979   123579: contig of 2601 bp in length
            *  123580 123679: gap of      100 bp
            *  123680   128211: contig of 4532 bp in length
            *  128212 128311: gap of      100 bp
            *  128312   132536: contig of 4225 bp in length
            *  132537 132636: gap of      100 bp
            *  132637   137226: contig of 4590 bp in length
            *  137227 137326: gap of      100 bp
            *  137327   142521: contig of 5195 bp in length
            *  142522 142621: gap of      100 bp
            *  142622   147070: contig of 4449 bp in length
            *  147071 147170: gap of      100 bp
            *  147171   151866: contig of 4696 bp in length
```

```
            *   151867 151966: gap of      100 bp
            *   151967   156914: contig of 4948 bp in length
            *   156915 157014: gap of      100 bp
            *   157015   162849: contig of 5835 bp in length
            *   162850 162949: gap of      100 bp
            *   162950   169655: contig of 6706 bp in length
            *   169656 169755: gap of      100 bp
            *   169756   174040: contig of 4285 bp in length
            *   174041 174140: gap of      100 bp
            *   174141   180979: contig of 6839 bp in length
            *   180980 181079: gap of      100 bp
            *   181080   190751: contig of 9672 bp in length
            *   190752 190851: gap of      100 bp
            *   190852   202044: contig of 11193 bp in length
            *   202045 202144: gap of      100 bp
            *   202145   211327: contig of 9183 bp in length
            *   211328 211427: gap of      100 bp
            *   211428   224152: contig of 12725 bp in length
            *   224153 224252: gap of      100 bp
            *   224253   234236: contig of 9984 bp in length
            *   234237 234336: gap of      100 bp
            *   234337   245499: contig of 11163 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-81I5"
                     /clone_lib="RPCI-11 Human Male BAC"

  Query Match             64.0%;  Score 16;  DB 64;  Length 245499;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db 109551 CAGGTGCAGGTCAGCA 109536


RESULT  44
AP002552
LOCUS       AP002552   262278 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 3/20.
ACCESSION   AP002552 BA000007
VERSION     AP002552.1  GI:13359995
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 262278)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .262278
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            159. .1091
                     /gene="ECs0538"
     CDS             159. .1091
                     /gene="ECs0538"
                     /note="similar to YBAS_ECOLI gi|1786693 percent identity
                     98 in 310 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative glutaminase"
                     /protein_id="BAB33961.1"
                     /db_xref="GI:13359996"
                     /translation="MLDANKLQQAVDQAYTQFHSLNGGQNADYIPFLANVPGQLAAVA
                     IVTSDGNVYSAGDSDYRFALESISKVCTLALALEDVGPQAVQDKIGADPTGLPFNSVI
                     ALELHGGKPLSPLVNAGAIATTSLINAENAEQRWQRILHIQQQLAGELVALSDEVNQS
                     EQTTNFHNRAIAWLLYSAGYLYCDAMEACDVYTRQCSTLINTVELATLGATLAAGGVN
                     PLTHKRVLQADNVPYILAEMMMEGLYGRSGDWAYRVGLPGKSGVGGGILAVVPGVMGI
                     AAFSPPLDEEGNSVRGQKMVASVAKQLGYNVFKG"
     gene            1094. .2386
                     /gene="ECs0539"
     CDS             1094. .2386
                     /gene="ECs0539"
                     /note="similar to YBAT_ECOLI gi|1786694 percent identity
                     99 in 430 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative amino acid/amine transport protein"
                     /protein_id="BAB33962.1"
                     /db_xref="GI:13359997"
                     /translation="MMNTEGNNGNKPLGLWNVVSIGIGAMVGAGIFALLGQAALLMEA
                     STWVAFAFGGIVAMFSGYAYARLGASYPSNGGIIDFFRRGLGNGVFSLALSLLYLLTL
                     AVSIAMVARAFGAYAVQFLHEGSQEEHLILLYALGIIAVMTLFNSLSNHAVGRLEVIL
                     VGIKMMILLLLIIAGVWSLQPAHISVSAPPSSGAFFSCIGITFLAYAGFGMMANAADK
                     VKDPQVIMPRAFLVAIGVTTLLYISLALVLLSDVSALELEKYADTAVAQAASPLLGHV
                     GYVIVVIGALLATASAINANLFAVFNIMDNMGSERELPKRMNKSLWRQSTWGNIIVVV
                     LIMLMTAALNLGSLASVASATFLICYLAVFVVAIRLRHDIHASLPILIVGTLVMLLVI
                     VGFIYSLWSQGSRALIWIIGAILLSLIVAMVMKRNKTV"
     gene            2724. .4079
```

```
                     /gene="ECs0540"
     CDS             2724. .4079
                     /gene="ECs0540"
                     /note="putative transporter (outer membrane
                     protein),similar to outer membrane transport proteins e.g.
                     CyaE protein [Bordetella pertussis]
                     gi|117799|sp|P11092|CYAE_BORPE percent identity 25 in 385
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative outer membrane transport protein"
                     /protein_id="BAB33963.1"
                     /db_xref="GI:13359998"
                     /translation="MARFQFKNRKNNGLIFFISFMVMGEAAIAAPLPQWANAPAVTPV
                     AQLSLQESILRAFARNPGVTQQAAQIGIGEAQIDEAKSAWYPHVGLTGNAGPSRQTDS
                     SGRLDNNVSYGITLTQLVYDFGKTNNDINLQTAARDSYRFKLMATLTDVAEKTATAYM
                     EVSRYQALCDAAQRNIHSLENVYNMAALRANAGLNSSSDELQAQTRIAGMRSTLEQYQ
                     AQMASAKAQLAVLTGVQPEAIAAPPAELAEQPVSLKNIDYQSIPLVLAAENLRQSAQY
                     GVEKTKAQYWPTLSIQGGKTRYQTSDRSYWDDQLQLNVNAPLYQGGAVSAQVQQAEGQ
                     QKISASQVEQAKLDVLQRASVAYANWTGARGREEAGLAQSESAHKTRDVYQNEYKLGK
                     RSLNDLLTVEQDVFQAQSAEINANYDGWVAAVNYAAAVNNLIPLAGIKQGLYNDLPDL
                     K"
     gene            4182. .8567
                     /gene="ECs0541"
     CDS             4182. .8567
                     /gene="ECs0541"
                     /note="unknown,similar to hypothetical proteins e.g.
                     [Synechocystis sp. strain PCC 6803] gi|7469433|pir||S76109
                     percent identity 33 in 1384 aa,also similar to RTX protein
                     [Aeromonas salmonicida]
                     gi|6752871|gb|AAF27914.1|AF218037_1 percent identity 33 in
                     138"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB33964.1"
                     /db_xref="GI:13359999"
                     /translation="MSLIIDVISRKTSVKQTLINPGDVTVVIYEPSVVQVHAQASAVA
                     RYVREGNDLLIYMQDGTVIRCNGYFLQAANTAEQSELVFADGQQLTHITFADTAAGGL
                     APVELTAQTTAIESIAPFLDTVAQTSAFPWGWLAGAAVGGGALGALLASGGDGDSKTE
                     VINNPTPPAEPGNATPSFLVTDNQGDQRGILATNDITDDTTPTFSGSGQAGATIQIKD
                     SNGNTIASTQVDNNGHWSVSLPTQSAGEHTWSVVQIVGSTITDAGSITLTIDNSQASV
                     QVATTAGDNIINASEQAAGFTLSGTSSHLAQGTELTVTLNGKTYTTSVGANGAWSVQV
                     PTADAQALGEGNQAVLVSGKDATGNTVTGAQLLTVDTQPPTLAINTIAQDNIISAAEH
                     NVALVLSGTSNAEAGQTVTLTVNGKSHTATVGSDGTWQVTLPATEVQALAEGNYAVNA
                     SVSDRAGNTTSHSANFTVDTSAPVVSVNTVAGDDILNNAEQAVAQIISGQVSGASPGD
                     TVTVKLGTHVLTGIVLADGSWNVALDPAVTRTLDRGANTIFVTVTDAAGNTGAASRAI
                     TLVGVSPLITINTVSGDDIISGAEKGAPLTLTGSTQQAETGQTVTVTLAGQSFTTTVQ
                     ADGSWSLTVPAAAMGNLPDGAVAITASVTDLSGNTGNTSRTITVDSQAPALSIDPLTA
                     DNIINAAESGQDLPITGTTDAQPGQTVTVTLNGQTYQGVVQPDGTWSVTVPAANVGAL
                     ADGNATVTASVNDVAGNPSSVSRVALVDATPPVVTINPVATDNVINTPEHAQAQIISG
                     TVTGAQAGDIVTVTLNNVDYTTVVDGSGNWSLGVPASVVSGLADGSYPVSVSVTDKAG
                     NTGSQSLTVTVNTAAPLIGINSIAGDDVINASEKGADLQITGTSDQPVNTAITVTLNG
                     QNYTTTTDASGNWSVTVPASAVTALGQANYTVTAAVTSDIGNSATASHNVLVDSALPG
                     VTINPVATDDIINAAEAGVAQTISGQVTGAEDGDTVTITLGGNTYTATVGSNLTWSVD
                     VPAADIQALGNGDLTVNASVTNQNGNTGSGTRDITIDANLPGLRVDTVAGDDVVNIIE
                     HGQALVVTGSSSGLAESTPLTVTINNVEYTTAVQADGSWSVGVTAAQVSAWPAGTVNI
                     AVSGESSAGNSVSITHPVTVDLTPAAITINTIATDDVINAAEKGADLTLSGTTTNVEP
                     GQTVTVTFGGKNYTASVASDGSWTATVPAADLASLPEGSASALASVSNINGNSASAVH
                     NYSVDSSAPTIIINTVASDNIVNASEADAGVTVSGSTTAEAGQIVTITLNSPTVQTYQ
                     ATVQADGSWSINIPAADLEALTDGSHTLTATVNDKAGNPASTTHNLAVDLTVPVLTIN
                     TIAGDDIINATEHGQALVISGSSTGGEAGDVVTVTLNSKTYTTTLDASGNWSVGVPAA
                     DVTALGSGPQTVTATVTDAAGNSDN"
     gene            8775. .24650
                     /gene="ECs0542"
     CDS             8775. .24650
                     /gene="ECs0542"
                     /note="unknown,similar to hypothetical proteins e.g.
                     [Synechocystis sp. strain PCC 6803] gi|7469433|pir||S76109
                     percent identity 36 in 2014 aa,similar to RTX protein
                     [Aeromonas salmonicida]
                     gi|6752871|gb|AAF27914.1|AF218037_1 percent identity 36 in
                     2051 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB33965.1"
                     /db_xref="GI:13360000"
                     /translation="MPASAVGALGEASYTVTANVTDSAGNSNSASHNVQVNTALPGVT
                     INPVATDDIINAAESGNAQTISGQVTGAAAGDTVTVTLGGKTYTATVQGNLSWSVDVP
                     AADIQAIGNGNLTVNASVTNGVGNTGSGSRDITIDANLPGLRVDTVAGDDVVNSIEHA
                     QALVITGSSSGLAAGAALTVVINTVTYAATVLADGTWSVGVPAADVSNWPAGTVNITV
                     SGTNTAGTTSTITHPVTVDLAAVAISINTVSGDDVINAAEKGADLTLSGSTSGVEVGQ
                     TVTVTFGGKTYTATVAGDGSWTTTVPAADLSVLRDGDATVQASVSTINGNTASATHAY
                     SVDATAPTLAINTIATDDILNAAEAGNPLTISGSSTAEAGQTVTVTLNGVTYSGSVQA
                     DGSWSVSLPTADLSNLTASQYTVSASVSDKAGNPASANHGLAVDLTVPVLTINTVSGD
                     DIINAAEHGQALVISGSSTGGEAGDVITVTLNSKTYTTMLDASGNWSVGVPAADVTAL
                     GSGPQTITAAITDAAGNSDDASRTVTVNLAAPTIGINTIATDDVIKATEKGADLQITG
                     TSNQPAGTTITVTLNGQNYTATTDSNGNWSATVPASAVSALGEANYTVTANVTDTAGN
                     SNSASHNVLVNSALPAVTINAVATDDIINAAESGNAQTISGQVTGAAQGDTVTVTLGG
                     NTYTATVQSNLSWSVDVPAADIQALGNGDLTVNASVTNGVGNTGSGSRDITIDANLPG
                     LRVDTVAGDDVINSIEHNQALVITGSSSGLTAGTALTVEINNVTYGATVLADGTWSLG
                     VPAVDVSNWPAGTVNITVSGTNSAGTTSTITHPVTVDLAGVAITINTLSGDDVINAVE

  Query Match              64.0%;  Score 16;  DB 2;  Length 262278;
  Best Local Similarity    100.0%;  Pred. No. 43;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcacgtt 24
          ||||||||||||||||
Db 152470 TGCAGGTCAGCACGTT 152485


RESULT  45
HSMX1B
LOCUS       HSMX1B     313064 bp    DNA             PRI       23-SEP-2000
DEFINITION  Homo sapiens chromosome 21 from 5 PACs and 5 Cosmids map
            21q22.2,D21S349-MX1; segment 2/2, complete sequence.
ACCESSION   AL442167 AJ011929
VERSION     AL442167.1  GI:10303260
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313064)
  AUTHORS   Ramser,J., Francis,F., Beck,A., Hennig,S., Klages,S., Borzym,K.,
            Langer,I., Steffens,C., Hildmann,T., Dagand,E., Yaspo,M.,
            Reinhardt,R. and Lehrach,H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 313064)
  AUTHORS   MPIMG.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) MPIMG, Abt. Lehrach, Max Planck Institut
            fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
COMMENT     Clones received from Resource Centre of the Human Genome Project at
            the  Max-Planck-Institut for Molecular Genetics.
            292889. .294336 sequence from clone KB447A5, accession no. AP001609
            (DDBJ), sequenced at KEIO university, Tokyo, Japan.
            Bases 1. .50 overlap with AL442166.
FEATURES             Location/Qualifiers
     source          1. .313064
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.2,D21S349-MX1"
                     /clone="PAC RPCI-1 247E2"
                     /clone="PAC RPCI-1 146B4"
                     /clone="PAC RPCI-1 141D16"
                     /clone="PAC RPCI-1 269A14"
                     /clone="PAC RPCI-1 265B9"
                     /clone="cosmid LLNLc116 44C5"
                     /clone="cosmid LLNLc116 16H18"
```

```
                     /clone="cosmid LLNLc116 14C10"
                     /clone="cosmid LLNLc116 25D2"
                     /clone="cosmid LLNLc116 87D5"
                     /clone_lib="LL21NC02, Chromosome 21 specific cosmid
                     library, Lawrence Livermore National Laboratory (LLNL),
                     creator: Pieter de Jong; PAC: RPCI1,3-5, Roswell Park
                     Cancer Institute, creator: Pieter de Jong, P.Ioannou"
     source          292889. .294336
                     /note="sequence from clone KB447A5, accession no. AP001609
                     (DDBJ), sequenced at KEIO university, Tokyo, Japan."
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
BASE COUNT    83213 a  71272 c  73589 g  84990 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 93;  Length 313064;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db 182813 GTCAGGTGCAGGTCAG 182828


Search completed: October  9, 2001, 15:49:28
Job time: 14954 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:53:33 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 34718

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	1041	20	AAX01303	E. coli biotin syn
2		25	100.0	1084	10	AAN91329	E.coli Bio B gene.
3		25	100.0	1121	7	AAN60496	Sequence encoding
4		25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	c 5	15	60.0	353	21	AAC56641	Eucalyptus grandis
	c 6	15	60.0	630	16	AAQ78948	Human immunoglobul
	c 7	15	60.0	4674	20	AAX90421	Human ataxin-2 lik
8		14	56.0	269	21	AAA31747	Plant microsatelli
	c 9	14	56.0	327	21	AAC00048	Human secreted pro
	c 10	14	56.0	374	22	AAF64981	Novel human polynu
	c 11	14	56.0	423	14	AAQ45597	Sequence encoding
	c 12	14	56.0	423	20	AAZ39427	Antibody B17X2 VH
	c 13	14	56.0	423	20	AAZ23965	Human B17X2 antibo
	c 14	14	56.0	444	18	AAT80181	Monoclonal antibod
	c 15	14	56.0	447	20	AAZ20406	IgG antibody 2.4.4
	c 16	14	56.0	613	16	AAQ78943	Human immunoglobul
	c 17	14	56.0	981	21	AAA15905	Human protein clon
	c 18	14	56.0	1100	21	AAA15915	Human protein clon
	c 19	14	56.0	1130	21	AAA39064	Human secreted pro
	c 20	14	56.0	1135	21	AAA39087	Human secreted pro
	c 21	14	56.0	1161	22	AAA91019	Human secreted pro
	c 22	14	56.0	1173	21	AAZ98136	Human signal pepti
	c 23	14	56.0	1186	21	AAC98064	Human colon cancer
24		14	56.0	1191	19	AAT98726	DNA encoding a S.
	c 25	14	56.0	1368	22	AAF44623	Novel protein kina
26		14	56.0	1634	19	AAZ96242	S. pneumoniae deri
27		14	56.0	1634	19	AAV37360	Streptococcus pneu
28		14	56.0	37856	21	AAA11992	S. cellulosum DNA
29		14	56.0	534720	19	AAV30458	Rhizobium species
30		14	56.0	536165	19	AAV30459	Rhizobium species
31		13	52.0	18	12	AAQ13647	Exon B/intron junc
	c 32	13	52.0	123	20	AAX85925	Oligonucleotide us
	c 33	13	52.0	225	14	AAQ60177	Human brain Expres
	c 34	13	52.0	276	21	AAC00049	Human secreted pro
	c 35	13	52.0	363	20	AAX85929	DNA encoding the h
	c 36	13	52.0	363	20	AAX79527	Heavy chain coding
37		13	52.0	375	21	AAF22004	Human breast and o
	c 38	13	52.0	378	18	AAT61237	Human anti-RSV mon
	c 39	13	52.0	378	18	AAT61239	Human anti-RSV mon
	c 40	13	52.0	400	21	AAC00033	Human secreted pro
	c 41	13	52.0	423	20	AAX79207	Anti-HIV-1 gp120 a
	c 42	13	52.0	437	21	AAZ42289	Human 5' EST isola
43		13	52.0	472	21	AAC75419	Human ORFX ORF974
	c 44	13	52.0	477	20	AAZ20408	IgG antibody 2.3.2
45		13	52.0	527	20	AAV87894	EST clone FG380.

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 tcgtcaggtgcaggtcagcacgttg 135


RESULT   2
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    134 tcgtcaggtgcaggtcagcacgttg 158


RESULT   3
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    152 tcgtcaggtgcaggtcagcacgttg 176


RESULT   4
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
```

```
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00024;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 tcgtcaggtgcaggtcagcacgttg 251


RESULT   5
AAC56641/c
ID   AAC56641 standard; DNA; 353 BP.
XX
AC   AAC56641;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #512.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
```

```
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 482; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;


  Query Match             60.0%;  Score 15;  DB 21;  Length 353;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggt 15
           ||||||||||||||
Db     326 TCGTCAGGTGCAGGT 312


RESULT   6
AAQ78948/c
ID   AAQ78948 standard; DNA; 630 BP.
XX
AC   AAQ78948;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human immunoglobulin Vh gene #10.
XX
KW   Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
KW   cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..495
FT                   /*tag=  a
FT                   /product= human immunoglobulin variable heavy chain
FT   intron          114..199
FT                   /*tag=  b
FT   misc_signal     322..324
FT                   /*tag=  c
FT                   /transl_except= unused termination codon
FT   misc_signal     373..375
FT                   /*tag=  d
FT                   /transl_except= unused termination codon
FT   misc_signal     496..498
FT                   /*tag=  e
FT                   /note= "miscellaneous signal, does not conform to
FT                          termination or splice site sequence"
XX
PN   WO9426895-A.
XX
PD   24-NOV-1994.
XX
PF   10-MAY-1993;   93WO-JP00603.
XX
PR   10-MAY-1993;   93WO-JP00603.
XX
PA   (NISB ) JAPAN TOBACCO INC.
XX
PI   Honjo T,  Matsuda F;
XX
DR   WPI; 1995-006791/01.
DR   P-PSDB; AAR66304.
XX
PT   DNA fragment comprising human immunoglobulin Vh genes - for the
PT   production of human immunoglobulin in mammalian hosts
XX
PS   Claim 19; Page 43-44; 130pp; Japanese.
XX
CC   A series of genes (AAQ78939-79002) encoding human immunoglobulin
CC   variable heavy chains. The genes were isolated and cloned from a series
CC   of cosmid constructs: Y202; Y103; Y21; Y6;Y24; 3-31; M84; M118 and M131,
CC   by PCR amplification using primers AAQ78917-38.  The genes are subdivided
CC   into 5 families of Vh genes.  The fragments cover a region of 800 kb.
CC   The DNA fragments were isolated from high molecular weight DNA from
CC   human placenta.  The DNA was partially digested with TaqI restriction
CC   enzyme. The fragments were separated by gel elecrophoresis and 35-45 kb
CC   fractions were collected.  The fragments were ligated with ClaI-digested
CC   cosmid vector pJB81.  The ligation products were in vitro packed and
CC   infected into E.coli 490A.  The fragments were then subcloned by colony
CC   hybridisation.  The Vh genes and the DNA fragments encoding them are
CC   useful in producing human immunoglobulin in mammalian hosts.
XX
SQ   Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 16;  Length 630;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagca 20
           ||||||||||||||
Db     280 AGGTGCAGGTCAGCA 266


RESULT   7
AAX90421/c
ID   AAX90421 standard; cDNA; 4674 BP.
XX
AC   AAX90421;
XX
DT   29-SEP-1999  (first entry)
XX
DE   Human ataxin-2 like gene.
XX
KW   Human; ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;
KW   cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..3218
```

FT /*tag= a
FT /product= "ataxin-2 like protein"
XX
PN WO9936527-A1.
XX
PD 22-JUL-1999.
XX
PF 19-JAN-1998; 98WO-CN00009.
XX
PR 19-JAN-1998; 98WO-CN00009.
XX
PA (UYHU-) UNIV HUNAN MEDICAL.
XX
PI Deng H, Liu C, Wang D, Xia J;
XX
DR WPI; 1999-458463/38.
DR P-PSDB; AAY29321.
XX
PT Ataxin-2 like protein, and related polynucleotides, useful in
PT treatment and diagnosis of ataxia,
XX
PS Claim 4; Page 21-23; 34pp; English.
XX
CC The present sequence represents a human ataxin-2 like gene. Ataxin-2
CC like polynucleotides (I), vectors containing (I) and recombinant host
CC cells are useful for recombinant production of Ataxin-2 like protein.
CC Both (I), ataxin-2 like protein and antibodies against ataxin-2 like
CC protein are useful as research reagents, for screening assays and in
CC diagnostic assays. Antagonists and agonists of ataxin-2 like protein
CC can be used to inhibit or enhance, respectively, the activity of
CC ataxin-2 like protein or expression of (I). Anti-ataxin-2 like protein
CC antibodies and ataxin-2 like protein or its fragments can be used
CC in vaccines. In particular, the proteins, antibodies, agonists and
CC antagonists can be used for treating, e.g. ataxia, cardiomyopathy,
CC deafness, neurological disease, cancer and AIDS, related to both an
CC excess and insufficient amounts of ataxin-2 like.
XX
SQ Sequence 4674 BP; 971 A; 1566 C; 1244 G; 893 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 4674;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgcaggtcagc 19
||||||||||||||||
Db 3741 CAGGTGCAGGTCAGC 3727

RESULT 8
AAA31747
ID AAA31747 standard; DNA; 269 BP.
XX
AC AAA31747;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #708.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 285; 392pp; English.
XX
CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 269 BP; 63 A; 62 C; 61 G; 83 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 269;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caggtcagcacgtt 24
||||||||||||||
Db 76 caggtcagcacgtt 89

RESULT 9
AAC00048/c
ID AAC00048 standard; cDNA; 327 BP.
XX
AC AAC00048;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 46.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00042.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 46; 71pp + CD-ROM; English.
XX

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other;

Query Match 56.0%; Score 14; DB 21; Length 327;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 144 AGGTGCAGGTCAGC 131

RESULT 10
AAF64981/c
ID AAF64981 standard; cDNA; 374 BP.
XX
AC AAF64981;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 737.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 650; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 374;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 180 AGGTGCAGGTCAGC 167

RESULT 11
AAQ45597/c
ID AAQ45597 standard; DNA; 423 BP.
XX
AC AAQ45597;
XX
DT 04-DEC-1993 (first entry)
XX
DE Sequence encoding the VH of antibody B17X2.
XX
KW Variable heavy antibody chain; human subgroup 4 germline; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 7..423
FT /*tag= a
XX
PN WO9312231-A.
XX
PD 24-JUN-1993.
XX
PF 13-DEC-1991; 91WO-AU00583.
XX
PR 13-DEC-1991; 91WO-AU00583.
XX
PA (DOWC) DOW CHEM AUSTRALIA LTD.
XX
PI Johnson KS, Mezes PS, Richard RA;
XX
DR WPI; 1993-214173/26.
DR P-PSDB; AAR38315.
XX
PT New composite antibody binding to tumour associated TAG-72
PT antigen - includes light chain variable region from human
PT subgroup 4 germline gene, useful, opt. as conjugate, for
PT diagnosis or treatment of cancer
XX
PS Disclosure; Figure 4; 150pp; English.
XX
CC Cell line B17X2 expresses an antibody utilising a variable light
CC chain encoded by a gene derived from Hum4 VL and a variable heavy
CC chain which makes a stable VL and VH combination.
XX
SQ Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 14; Length 423;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    133 AGGTGCAGGTCAGC 120


RESULT  12
AAZ39427/c
ID   AAZ39427 standard; DNA; 423 BP.
XX
AC   AAZ39427;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Antibody B17X2 VH nucleotide sequence.
XX
KW   Tumor associated sialylated glycoprotein; TAG-72; cancer antigen;
KW   carcinoma lesion; diagnostic; cancer; antibody; human; B17X2;
KW   anti-mouse antibody hypersensitivity reaction; ss.
XX
OS   Homo sapiens.
XX
PN   US5976845-A.
XX
PD   02-NOV-1999.
XX
PF   07-JUN-1995;   95US-0487743.
XX
PR   16-JUN-1994;   94US-0261354.
PR   19-APR-1990;   90US-0510697.
PR   20-OCT-1992;   92US-0964536.
XX
PA   (DOWC ) DOW CHEM CO.
XX
PI   Johnson KS,  Richard RA,  Mezes PS;
XX
DR   WPI; 1999-619651/53.
DR   P-PSDB; AAY57179.
XX
PT   Production of humanized anti-TAG-72 antibodies, used for the detection,
PT   in vivo imaging and treatment of cancers  -
XX
PS   Disclosure; Fig 4A-B; 85pp; English.
XX
CC   The invention relates to producing humanized anti-tumor associated
CC   sialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies
CC   have binding specificity for the cancer antigen TAG-72. These antibodies
CC   have variable regions with VL segments derived from human subgroup IV
CC   germline gene and a VH segment (encoded by the VHalphaTAG germline gene)
CC   which is capapble of combining with the VL to form a threee dimensional
CC   structure having the ability to bind TAG-72. They can be used for the in
CC   vivo detection of carcinoma lesions. They can also be used for in vitro
CC   diagnostics. They can also be modified with therapeutic agents e.g. a
CC   radionuclide, drug, biological response modifier, toxin or another
CC   antibody for the treatment of cancers. The humanized anti-TAG-72
CC   antibodies can reduce harmful anti-mouse antibody hypersensitivity
CC   reactions.
XX
SQ   Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    133 AGGTGCAGGTCAGC 120


RESULT  13
AAZ23965/c
ID   AAZ23965 standard; DNA; 423 BP.
XX
AC   AAZ23965;
XX
DT   09-FEB-2000  (first entry)
XX
DE   Human B17X2 antibody VH segment DNA.
XX
KW   Human; antibody; humanized; anti-tumor; sialylated glycoprotein antibody;
KW   TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion;
KW   diagnostic; treatment; ss.
XX
OS   Homo sapiens.
XX
PN   US5976531-A.
XX
PD   02-NOV-1999.
XX
PF   16-JUN-1994;   94US-0261354.
XX
PR   19-APR-1990;   90US-0510697.
PR   20-OCT-1992;   92US-0964536.
XX
PA   (DOWC ) DOW CHEM CO.
XX
PI   Johnson KS,  Mezes PS,  Richard RA;
XX
DR   WPI; 1999-632731/54.
DR   P-PSDB; AAY50688.
XX
PT   New humanized anti-TAG-72 antibodies, used for the detection, in vivo
PT   imaging and treatment of cancers  -
XX
PS   Disclosure; Figure 4A-B; 83pp; English.
XX
CC   This invention describes novel humanized anti-tumor associated
CC   sialylated glycoprotein antibodies (TAG-72) which have cytostatic
CC   activity. The antibodies have binding specificity for the cancer antigen
CC   TAG-72.  They can be used for the in vivo detection of carcinoma lesions.
CC   They can also be used for in vitro diagnostics. They can also be modified
CC   with therapeutic agents e.g. a radionuclide, drug, biological response
CC   modifier, toxin or another antibody for the treatment of cancers. The
CC   humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody
CC   hypersensitivity reactions.
XX
SQ   Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    133 AGGTGCAGGTCAGC 120


RESULT  14
AAT80181/c
ID   AAT80181 standard; DNA; 444 BP.
XX
AC   AAT80181;
XX
DT   02-APR-1998  (first entry)
XX
DE   Monoclonal antibody 105AD7 heavy chain variable region DNA sequence.
XX
KW   Monoclonal antibody 105AD7; complementarity determining region;
KW   CDR; human; immune response; treatment; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..444
FT                   /*tag=  a
```

```
FT                   /transl_except= (pos: 157..159, aa: Gly)
FT                   /transl_except= (pos: 160..162, aa: Val)
FT                   /product= "105AD7 heavy chain variable region"
XX
PN   WO9732021-A1.
XX
PD   04-SEP-1997.
XX
PF   28-FEB-1997;   97WO-GB00591.
XX
PR   29-FEB-1996;   96GB-0004321.
PR   28-FEB-1996;   96GB-0004177.
XX
PA   (CANC-) CANCER RES CAMPAIGN TECHNOLOGY LTD.
XX
PI   Durrant LG,  Robins RA,  Spendlove I;
XX
DR   WPI; 1997-448689/41.
DR   P-PSDB; AAW26239.
XX
PT   Nucleic acid encoding peptide derived from antibody 105AD7 - used to
PT   induce an immune response for treatment and prevention of cancer
XX
PS   Claim 1; Fig 1A; 32pp; English.
XX
CC   This is the DNA sequence of the monoclonal antibody 105AD7 heavy chain
CC   variable region. 105AD7 is a human monoclonal anti-idiotypic antibody
CC   that mimics T cell epitopes on the tumour associated antigen gp72 and
CC   has been used for therapeutic vaccination. A peptide derived from this
CC   antibody contains the 105AD7 heavy chain variable region and a 105AD7
CC   kappa chain. The complementarity determining regions (CDR) of this
CC   peptide and its fragments can stimulate immune responses. The peptide
CC   can be used for screening for its functional equivalents and mimetics.
CC   The peptides, and the fragments, equivalents and mimetics are used to
CC   stimulate an immune response to a tumour antigen, for the treatment or
CC   prevention of tumours. Transfected host cells are used to produce this
CC   peptide and the vectors can be used to generate the peptide in vivo for
CC   stimulation of an immune response. The peptide contains promiscuous
CC   helper epitopes which stimulate a response against tumour antigens other
CC   than gp72. These epitopes may help a cytotoxic T lymphocyte response to
CC   any co-injected antigen.
XX
SQ   Sequence 444 BP; 100 A; 134 C; 107 G; 103 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    121 AGGTGCAGGTCAGC 108


RESULT  15
AAZ20406/c
ID   AAZ20406 standard; cDNA; 447 BP.
XX
AC   AAZ20406;
XX
DT   19-NOV-1999  (first entry)
XX
DE   IgG antibody 2.4.4 heavy chain coding sequence.
XX
KW   Antibody; CD147; IgM; ABX-CBL; activated T-cell killing; leukaemia;
KW   activated B-cell; monocyte; graft versus host disease; therapy; cancer;
KW   organ transplant rejection disease; lymphoma; pancreatic disease;
KW   autoimmune disease; inflammatory disease; arthritis; binding site; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..447
FT                   /*tag=  a
FT                   /note= "no stop codon given"
XX
PN   WO9945031-A2.
XX
PD   10-SEP-1999.
XX
PF   03-MAR-1999;   99WO-US04583.
XX
PR   03-MAR-1998;   98US-0034607.
PR   03-FEB-1999;   99US-0244253.
XX
PA   (ABGE-) ABGENIX INC.
XX
PI   Davis CG,  Blacher RW,  Corvalan JR,  Culwell AR,  Green LL,  Hales J;
PI   Havrilla N,  Ivanov VE,  Lipani JA,  Liu Q,  Weber RF,  Yang X;
XX
DR   WPI; 1999-540816/45.
DR   P-PSDB; AAY34305.
XX
PT   New monoclonal antibody, used for treating e.g. graft versus host
PT   disease, cancers, autoimmune diseases and inflammatory diseases  -
XX
PS   Disclosure; Fig 30; 245pp; English.
XX
CC   This sequence encodes the heavy chain of an antibody of the
CC   invention. The antibody is a monoclonal antibody (MAb) with an isotype
CC   that fixes complement and a variable region that binds to the epitope on
CC   CD147 bound by the IgM MAb ABX-CBL, providing that the antibody is not
CC   CBL1. The MAb can selectively kill activated T-cells, activated B-cells
CC   or resting or activated monocytes. The products and methods can be used
CC   for treating diseases involving activated T-cells or B-cells or
CC   monocytes, e.g. graft versus host disease (GVHD), organ transplant
CC   rejection diseases (e.g. renal transplant, ocular transplant), cancers
CC   (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
CC   pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
CC   (e.g. arthritis).
XX
SQ   Sequence 447 BP; 96 A; 150 C; 106 G; 95 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 447;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db     34 AGGTGCAGGTCAGC 21


RESULT  16
AAQ78943/c
ID   AAQ78943 standard; DNA; 613 BP.
XX
AC   AAQ78943;
XX
DT   07-AUG-1995  (first entry)
XX
DE   Human immunoglobulin variable heavy chain #5.
XX
KW   Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
KW   cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..513
FT                   /*tag=  a
FT                   /product= human immunoglobulin variable heavy chain
FT   intron          117..202
FT                   /*tag=  b
FT   misc_signal     514..516
FT                   /*tag=  c
```

```
FT                   /note= "miscellaneous signal, does not conform to
FT                         terminator or splice site sequence"
XX
PN   WO9426895-A.
XX
PD   24-NOV-1994.
XX
PF   10-MAY-1993;   93WO-JP00603.
XX
PR   10-MAY-1993;   93WO-JP00603.
XX
PA   (NISB ) JAPAN TOBACCO INC.
XX
PI   Honjo T,  Matsuda F;
XX
DR   WPI; 1995-006791/01.
DR   P-PSDB; AAR66299.
XX
PT   DNA fragment comprising human immunoglobulin Vh genes - for the
PT   production of human immunoglobulin in mammalian hosts
XX
PS   Disclosure; Page 36-37; 130pp; Japanese.
XX
CC   A series of genes (AAQ78939-79002) encoding human immunoglobulin
CC   variable heavy chains. The genes were isolated and cloned from a series
CC   of cosmid constructs: Y202; Y103; Y21; Y6;Y24; 3-31; M84; M118 and M131,
CC   by PCR amplification using primers AAQ78917-38.  The genes are subdivided
CC   into 5 families of Vh genes.  The fragments cover a region of 800 kb.
CC   The DNA fragments were isolated from high molecular weight DNA from
CC   human placenta.  The DNA was partially digested with TaqI restriction
CC   enzyme. The fragments were separated by gel elecrophoresis and 35-45 kb
CC   fractions were collected.  The fragments were ligated with ClaI-digested
CC   cosmid vector pJB81.  The ligation products were in vitro packed and
CC   infected into E.coli 490A.  The fragments were then subcloned by colony
CC   hybridisation.  The Vh genes and the DNA fragments encoding them are
CC   useful in producing human immunoglobulin in mammalian hosts.
XX
SQ   Sequence 613 BP; 139 A; 194 C; 139 G; 141 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 16;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    283 AGGTGCAGGTCAGC 270


RESULT  17
AAA15905/c
ID   AAA15905 standard; cDNA; 981 BP.
XX
AC   AAA15905;
XX
DT   12-JUN-2000  (first entry)
XX
DE   Human protein clone HP10568 coding sequence.
XX
KW   Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW   cytokine production; cell proliferation; cell differentiation;
KW   immune deficiency; infectious disease; autoimmune disorder; asthma;
KW   multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW   allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW   nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW   systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW   coagulation disorder; myocardial infarction; inflammatory condition;
KW   septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW   nephritis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200005367-A2.
XX
PD   03-FEB-2000.
XX
PF   22-JUL-1999;   99WO-JP03929.
XX
PR   24-JUL-1998;   98JP-0208820.
PR   07-AUG-1998;   98JP-0224105.
PR   25-AUG-1998;   98JP-0238116.
PR   09-SEP-1998;   98JP-0254736.
PR   29-SEP-1998;   98JP-0275505.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2000-182694/16.
DR   P-PSDB; AAY94857.
XX
PT   Novel human proteins having hydrophobic domains useful for treating
PT   osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT   multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS   Claim 3; Page 189; 351pp; English.
XX
CC   This sequence encodes a human protein of the invention, which has
CC   hydrophobic domains. The DNA sequences can be used as a probe or as a
CC   genetic marker. The protein can also be used as a marker, and to identify
CC   potential genetic disorders. The DNA and protein can also be used as
CC   nutritional sources or supplements. The protein exhibits cytokine, cell
CC   proliferation, cell differentiation activities and induces production of
CC   other cytokines in certain cell populations. The protein also exhibits
CC   immune stimulating or immune suppressing activity. It can be used in the
CC   treatment of various immune deficiencies and disorders, and to treat
CC   infectious diseases caused by viral, bacterial, fungal or other
CC   infections. The protein is also used for treating autoimmune disorders
CC   such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC   arthritis. It is also useful in the treatment of allergic reactions and
CC   conditions such as asthma, and in immune suppression after organ
CC   transplantation. The protein is useful in regulation of haematopoiesis
CC   and consequently in the treatment of myeloid or lymphoid cell
CC   deficiencies. It is also used in compositions for tissue growth or
CC   regeneration. The protein is also used in the treatment of osteoporosis
CC   or osteoarthritis and in the treatment of periodontal disease and other
CC   tooth repair processes. The protein is used in the treatment of nervous
CC   system disorders such as Alzheimer's disease, Parkinson's disease, and
CC   Huntington's disease. They are useful for protection or regeneration and
CC   treatment of lung or liver fibrosis, reperfusion injury in various
CC   tissues, and conditions resulting from systemic cytokine damage. They are
CC   also used for promoting or inhibiting tissue differentiation. They are
CC   also used as contraceptives since they exhibit activin or inhibin related
CC   activities and as a fertility inducing therapeutic. They are used for
CC   treating various coagulation disorders and in treatment and prevention of
CC   conditions resulting from coagulation activities e.g. myocardial
CC   infarction or stroke. They also acts as receptors, receptor ligands or
CC   inhibitors or agonists of receptor/ligand interactions. They are used to
CC   treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC   reperfusion injury, arthritis, and nephritis. They can be used to
CC   prevent tumours.
XX
SQ   Sequence 981 BP; 194 A; 299 C; 275 G; 213 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 981;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    139 AGGTGCAGGTCAGC 126
```

RESULT 18
AAA15915/c
ID AAA15915 standard; cDNA; 1100 BP.
XX
AC AAA15915;
XX
DT 12-JUN-2000 (first entry)
XX
DE Human protein clone HP10568 full length coding sequence.
XX
KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200005367-A2.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-JP03929.
XX
PR 24-JUL-1998; 98JP-0208820.
PR 07-AUG-1998; 98JP-0224105.
PR 25-AUG-1998; 98JP-0238116.
PR 09-SEP-1998; 98JP-0254736.
PR 29-SEP-1998; 98JP-0275505.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
XX
PI Kato S, Kimura T;
XX
DR WPI; 2000-182694/16.
DR P-PSDB; AAY94857.
XX
PT Novel human proteins having hydrophobic domains useful for treating
PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS Claim 4; Page 203-205; 351pp; English.
XX
CC This sequence encodes a human protein of the invention, which has
CC hydrophobic domains. The DNA sequences can be used as a probe or as a
CC genetic marker. The protein can also be used as a marker, and to identify
CC potential genetic disorders. The DNA and protein can also be used as
CC nutritional sources or supplements. The protein exhibits cytokine, cell
CC proliferation, cell differentiation activities and induces production of
CC other cytokines in certain cell populations. The protein also exhibits
CC immune stimulating or immune suppressing activity. It can be used in the
CC treatment of various immune deficiencies and disorders, and to treat
CC infectious diseases caused by viral, bacterial, fungal or other
CC infections. The protein is also used for treating autoimmune disorders
CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC arthritis. It is also useful in the treatment of allergic reactions and
CC conditions such as asthma, and in immune suppression after organ
CC transplantation. The protein is useful in regulation of haematopoiesis
CC and consequently in the treatment of myeloid or lymphoid cell
CC deficiencies. It is also used in compositions for tissue growth or
CC regeneration. The protein is also used in the treatment of osteoporosis
CC or osteoarthritis and in the treatment of periodontal disease and other
CC tooth repair processes. The protein is used in the treatment of nervous
CC system disorders such as Alzheimer's disease, Parkinson's disease, and
CC Huntington's disease. They are useful for protection or regeneration and
CC treatment of lung or liver fibrosis, reperfusion injury in various
CC tissues, and conditions resulting from systemic cytokine damage. They are
CC also used for promoting or inhibiting tissue differentiation. They are
CC also used as contraceptives since they exhibit activin or inhibin related
CC activities and as a fertility inducing therapeutic. They are used for
CC treating various coagulation disorders and in treatment and prevention of
CC conditions resulting from coagulation activities e.g. myocardial
CC infarction or stroke. They also acts as receptors, receptor ligands or
CC inhibitors or agonists of receptor/ligand interactions. They are used to
CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC reperfusion injury, arthritis, and nephritis. They can be used to
CC prevent tumours.
XX
SQ Sequence 1100 BP; 221 A; 333 C; 314 G; 232 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1100;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 195 AGGTGCAGGTCAGC 182

RESULT 19
AAA39064/c
ID AAA39064 standard; cDNA; 1130 BP.
XX
AC AAA39064;
XX
DT 30-AUG-2000 (first entry)
XX
DE Human secreted protein gene 13 SEQ ID NO:23.
XX
KW Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW hyperproliferative disorder; infectious disease; tissue regeneration;
KW screening; food additive; preservative; wound healing;
KW hyper-vascular disease; chromosome 11; ss.
XX
OS Homo sapiens.
XX
PN WO200017222-A1.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22012.
XX
PR 23-SEP-1998; 98US-0101546.
PR 02-OCT-1998; 98US-0102895.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;
XX
DR WPI; 2000-283538/24.
DR P-PSDB; AAB08903.
XX
PT Human secreted proteins and coding sequences useful in diagnostic and
PT therapeutic methods for disorders such as immune system or
PT proliferative disorders, related to the proteins -
XX
PS Claim 1; Page 329-330; 416pp; English.
XX
CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC secreted proteins and their related polynucleotide sequences are useful
CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and

```
CC   polynucleotide sequences may be useful for treating disorders of the
CC   immune system, hyperproliferative disorders, infectious disease,
CC   regeneration of tissues, for chemotaxis and for screening molecules that
CC   bind to the proteins. The proteins or polynucleotide sequences may be
CC   used as food additives or preservatives, to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, co-factors or other nutritional components. Agonists or
CC   antagonists of the proteins may be used to prevent scar tissue growth
CC   during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC   and AAB08890 are sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    201 AGGTGCAGGTCAGC 188


RESULT  20
AAA39087/c
ID   AAA39087 standard; cDNA; 1135 BP.
XX
AC   AAA39087;
XX
DT   30-AUG-2000  (first entry)
XX
DE   Human secreted protein gene 13 SEQ ID NO:46.
XX
KW   Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
KW   immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW   hyperproliferative disorder; infectious disease; tissue regeneration;
KW   screening; food additive; preservative; wound healing;
KW   hyper-vascular disease; chromosome 11; ss.
XX
OS   Homo sapiens.
XX
PN   WO200017222-A1.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22012.
XX
PR   23-SEP-1998;   98US-0101546.
PR   02-OCT-1998;   98US-0102895.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Rosen CA,  Duan RD,  Shi Y,  Lafleur DW,  Young PE,  Ni J;
PI   Komatsoulis G,  Endress GA,  Soppet DR;
XX
DR   WPI; 2000-283538/24.
DR   P-PSDB; AAB08926.
XX
PT   Human secreted proteins and coding sequences useful in diagnostic and
PT   therapeutic methods for disorders such as immune system or
PT   proliferative disorders, related to the proteins  -
XX
PS   Claim 1; Page 347-348; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC   human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC   secreted proteins and their related polynucleotide sequences are useful
CC   for diagnostic and therapeutic methods useful for diagnosing and treating
CC   disorders related to the secreted proteins. The proteins, and
CC   polynucleotide sequences may be useful for treating disorders of the
CC   immune system, hyperproliferative disorders, infectious disease,
CC   regeneration of tissues, for chemotaxis and for screening molecules that
CC   bind to the proteins. The proteins or polynucleotide sequences may be
CC   used as food additives or preservatives, to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, co-factors or other nutritional components. Agonists or
CC   antagonists of the proteins may be used to prevent scar tissue growth
CC   during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC   and AAB08890 are sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1135;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    188 AGGTGCAGGTCAGC 175


RESULT  21
AAA91019/c
ID   AAA91019 standard; DNA; 1161 BP.
XX
AC   AAA91019;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human secreted protein PRO7154 coding sequence.
XX
KW   Secreted protein; human; PRO protein; neoplastic cell growth; tumour;
KW   proliferation; leukaemia; lymphoid malignancy; inflammatory disorder;
KW   angiogenic disorder; immunologic disorder; PRO7154; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             102..1085
FT                   /*tag=  a
FT                   /product= PRO7154
XX
PN   WO200075317-A2.
XX
PD   14-DEC-2000.
XX
PF   15-MAY-2000; 2000WO-US13358.
XX
PR   09-JUN-1999;   99US-0138385.
PR   20-JUL-1999;   99US-0144790.
PR   03-AUG-1999;   99US-0146843.
PR   10-AUG-1999;   99US-0148188.
PR   17-AUG-1999;   99US-0149320.
PR   17-AUG-1999;   99US-0149327.
PR   17-AUG-1999;   99US-0149396.
PR   20-AUG-1999;   99US-0150114.
PR   31-AUG-1999;   99US-0151700.
PR   31-AUG-1999;   99US-0151734.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Botstein DA,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-071075/08.
DR   P-PSDB; AAY97585.
XX
PT   Antibodies against PRO polypeptides, useful for diagnosing and treating
PT   tumours are associated with gene amplification, neoplastic cell growth
PT   and proliferation in mammals -
XX
```

PS Claim 50; Fig 11; 143pp; English.
XX
CC This sequence encodes a human PRO protein of the invention. The PRO
CC proteins are secreted proteins. Antagonists or antibodies of PRO
CC polypeptides are useful for diagnosing and treating tumours are
CC associated with gene amplification, neoplastic cell growth and
CC proliferation in mammals, and those conditions characterised by
CC overexpression and/or activation of the amplified genes. Such conditions
CC include benign or malignant tumours (e.g. renal, liver, kidney, bladder,
CC breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval,
CC thyroid, hepatic carcinomas, sarcomas, glioblastomas and various head and
CC neck tumours); leukaemias and lymphoid malignancies; neuronal, glial,
CC astrocytal, hypothalamic, and other glandular, macrophageal, epithelial,
CC stromal and blastocoelic disorders; and inflammatory, angiogenic and
CC immunologic disorders. These may further be used to qualitatively or
CC quantitatively detect the expression of proteins encoded by the
CC amplified genes, and in tumour diagnostics or prognostics. The PRO
CC polypeptide or its antagonist may be used for the preparation of a
CC medicament in the treatment of a condition, which is responsive to the
CC PRO polypeptide, its antagonist or anti-PRO antibody.
XX
SQ Sequence 1161 BP; 235 A; 356 C; 327 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 1161;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 240 AGGTGCAGGTCAGC 227

RESULT 22
AAZ98136/c
ID AAZ98136 standard; cDNA; 1173 BP.
XX
AC AAZ98136;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-28 cDNA SEQ ID NO:162.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
PN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14484.
XX
PR 26-JUN-1998; 98US-0090762.
PR 31-JUL-1998; 98US-0094983.
PR 01-OCT-1998; 98US-0102686.
PR 11-DEC-1998; 98US-0112129.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX
DR WPI; 2000-160673/14.
DR P-PSDB; AAY87251.
XX
PT New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and
PT cardiovascular disease
XX
PS Claim 9; Page 268; 327pp; English.
XX
CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can
CC be used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense,
CC triplex-forming or ribozyme therapeutics, for detecting related sequences
CC or genetic variations, and for chromosomal mapping. HSPP are also used to
CC raise specific antibodies (Ab) and to screen for agonists and
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic
CC antagonists, in competitive drug screens, and for purification of HSPP
CC from natural sources.
XX
SQ Sequence 1173 BP; 236 A; 363 C; 330 G; 244 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1173;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 256 AGGTGCAGGTCAGC 243

RESULT 23
AAC98064/c
ID AAC98064 standard; cDNA; 1186 BP.
XX
AC AAC98064;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:74.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR P-PSDB; AAB53307.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 1; Page 520-521; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 1186 BP; 244 A; 361 C; 338 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 1186;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    255 AGGTGCAGGTCAGC 242
```

RESULT 24
AAT98726
ID AAT98726 standard; DNA; 1191 BP.
XX
AC AAT98726;
XX
DT 09-NOV-1998 (first entry)
XX
DE DNA encoding a S. pneumoniae protein of unknown function.
XX
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 596..979
FT /*tag= a
XX
PN WO9743303-A1.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US07950.
XX
PR 14-MAY-1996; 96US-0017670.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-008793/01.
DR P-PSDB; AAW38682.
XX
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 4; Pages 222-223; 483pp; English.
XX
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
XX
SQ Sequence 1191 BP; 294 A; 258 C; 294 G; 345 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1191;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 cgtcaggtgcaggt 15
          ||||||||||||||
Db    954 cgtcaggtgcaggt 967
```

RESULT 25
AAF44623/c
ID AAF44623 standard; cDNA; 1368 BP.
XX
AC AAF44623;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase cDNA, SEQ ID NO: 2.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
OS Mus musculus.
XX
PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGE-) SUGEN INC.

XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
DR P-PSDB; AAB65598.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Disclosure; Fig 2; 310pp; English.
XX
CC The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 1368 BP; 346 A; 359 C; 398 G; 265 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1368;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 1022 AGGTGCAGGTCAGC 1009

RESULT 26
AAZ96242
ID AAZ96242 standard; DNA; 1634 BP.
XX
AC AAZ96242;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #70.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-159452/14.
DR P-PSDB; AAY85867.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
XX
PS Claim 4; Page 101-102; 640pp; English.
XX
CC This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC AAY85792-Y86182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
SQ Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1634;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggt 15
||||||||||||||
Db 954 cgtcaggtgcaggt 967

RESULT 27
AAV37360
ID AAV37360 standard; DNA; 1634 BP.
XX
AC AAV37360;
XX
DT 13-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae coding region.
XX
KW coding region; ORF; open reading frame; antibacterial;
KW infection; prevention; meningitis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 1040..1291
FT /*tag= a
FT /product= unknown
XX
PN WO9819689-A1.
XX
PD 14-MAY-1998.
XX
PF 27-OCT-1997; 97WO-US19226.
XX
PR 01-NOV-1996; 96US-0029930.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX
DR WPI; 1998-286586/25.
DR P-PSDB; AAW60963.
XX
PT New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT e.g. for identifying anti-bacterial(s) for treatment and prevention
PT of meningitis
XX

```
PS   Claim 1; Page 60; 130pp; English.
XX
CC   The sequence is that of a coding region isolated from
CC   S. pneumoniae. Its encoded protein, or agonists of it,
CC   may be useful as an antibacterial for treatment or
CC   prevention of infection, specifically caused by S. pneumoniae
CC   (particularly meningitis) but possibly also Helicobacter
CC   pylori (ulcers and gastric cancer). It may be of particular
CC   use before insertion of an in-dwelling device or any other
CC   invasive procedure. The protein, or nucleic acid encoding
CC   it, can also be used in vaccines to induce a cellular
CC   and/or humoral immune response, or to screen for other
CC   antibacterials. The DNA may also contain flanking sequences
CC   that are potential sources of control elements for bacterial
CC   gene expression. Detecting a sequence encoding the protein
CC   can be used diagnostically, e.g. to detect a mutation for
CC   serotyping or classifying infectious agents.
XX
SQ   Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 19;  Length 1634;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggt 15
           ||||||||||||||
Db     954 cgtcaggtgcaggt 967


RESULT  28
AAA11992
ID   AAA11992 standard; DNA; 37856 BP.
XX
AC   AAA11992;
XX
DT   07-AUG-2000  (first entry)
XX
DE   S. cellulosum DNA encoding polyketide and hereropolyketide enzymes.
XX
KW   Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
KW   epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
KW   plant-protection; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (3398..6100)
FT                   /*tag=  a
FT                   /product= "ORF1-tRNA synthetase"
FT                   /note= "gtg start codon"
FT   CDS             complement (6374..7111)
FT                   /*tag=  b
FT                   /product= "ORF2-monooxygenase"
FT   CDS             complement (8433..9550)
FT                   /*tag=  c
FT                   /product= "ORF3-aminotransferase"
FT                   /note= "AGT start codon given in the specification"
FT   CDS             9855..11393
FT                   /*tag=  d
FT                   /product= "ORF4- tyrosine/DOPA-Decarboxylase"
FT                   /note= "GTG start codon"
FT   CDS             12212..13658
FT                   /*tag=  e
FT                   /product= "ORF5-3-oxoacyl-ACP-reductase"
FT                   /note= "ACC start codon"
FT   CDS             15374..19984
FT                   /*tag=  f
FT                   /product= "ORF6-polyketide synthase"
FT   CDS             20003..27889
FT                   /*tag=  g
FT                   /product= "ORF7-peptide synthetase"
FT   CDS             28251..29400
FT                   /*tag=  h
FT                   /product= "ORF8-transpeptidase"
FT   CDS             complement (30040..31720)
FT                   /*tag=  i
FT                   /product= "ORF9-regulation element"
FT                   /note= "CGC stop codon"
FT   CDS             31982..32932
FT                   /*tag=  j
FT                   /product= "ORF10-transcription regulator"
FT   CDS             33128..33613
FT                   /*tag=  k
FT                   /product= "ORF11-regulation element"
FT                   /note= "GTG start codon"
FT   CDS             33661..34077
FT                   /*tag=  l
FT                   /product= "ORF12-regulation element"
FT   CDS             complement (35255..35616)
FT                   /*tag=  m
FT                   /product= "ORF13-transcription regulator"
FT   CDS             complement (35730..36242)
FT                   /*tag=  n
FT                   /product= "ORF14-transcription regulator"
FT                   /note= "GTG start codon"
XX
PN   DE19846493-A1.
XX
PD   13-APR-2000.
XX
PF   09-OCT-1998;   98DE-1046493.
XX
PR   09-OCT-1998;   98DE-1046493.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Beyer S,  Mueller R;
XX
DR   WPI; 2000-294101/26.
XX
PT   DNA sequence coding for products involved in the biosynthesis of
PT   polyketide or heteropolyketide compounds, especially epothilone  -
XX
PS   Claim 3; Page 20-33; 36pp; German.
XX
CC   This invention describes a novel DNA sequence (I) whose expression
CC   products effect or are involved in the enzymatic biosynthesis,
CC   mutasynthesis or partial synthesis of polyketide or heteropolyketide
CC   compounds (II). (I) can be inserted into an expression vector and used
CC   to transform or transfect prokaryotic or eukaryotic cells with the aim
CC   of obtaining strains that produce large amounts of polyketide or
CC   heteropolyketide compounds, especially epothilones, which have cytotoxic
CC   and/or immunosuppressant and antibiotic and antifungal activities and
CC   are useful as plant-protection agents. This sequence represents the DNA
CC   sequence isolated from Sorangium cellulosum which is described in the
CC   method of the invention.
XX
SQ   Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 37856;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtc 16
           ||||||||||||||
Db   35785 gtcaggtgcaggtc 35798


RESULT  29
AAV30458
ID   AAV30458 standard; DNA; 534720 BP.
XX
AC   AAV30458;
XX
```

```
DT   14-OCT-1998  (first entry)
XX
DE   Rhizobium species plasmid pNGR234a.
XX
KW   Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW   degradation; metabolism; host range; nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             417796..418671
FT                   /*tag= a
FT                   /standard_name= "ORF K1"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppC gene"
FT   CDS             418673..419680
FT                   /*tag= b
FT                   /standard_name= "ORF K2"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppD gene"
FT   CDS             419677..420738
FT                   /*tag= c
FT                   /standard_name= "ORF K3"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppF gene"
FT   CDS             420774..422159
FT                   /*tag= d
FT                   /standard_name= "ORF K4"
FT                   /product= "encapsulation-like protein"
FT                   /note= "homologous to the CapA gene"
FT   CDS             422628..424031
FT                   /*tag= e
FT                   /standard_name= "ORF K5"
FT                   /product= "aminotransferase-like protein"
FT                   /note= "homologous to the BioA gene"
FT   CDS             424056..425594
FT                   /*tag= f
FT                   /standard_name= "ORF K6"
FT                   /product= "(semi)aldehyde dehydrogenase-like protein"
FT   CDS             complement (426949..428028)
FT                   /*tag= g
FT                   /standard_name= "ORF K7"
FT                   /product= "transposase homologue"
FT                   /note= "homologous to the Tnp gene"
FT   CDS             428292..429623
FT                   /*tag= h
FT                   /standard_name= "ORF K8"
FT                   /product= "glutamate dehydrogenase-like protein"
FT                   /note= "homologous to the GLUD1 gene"
FT   CDS             complement (430538..431284)
FT                   /*tag= i
FT                   /standard_name= "ORF K9"
FT                   /product= "transposase homologue"
FT   CDS             complement (431296..432840)
FT                   /*tag= j
FT                   /standard_name= "ORF K10"
FT                   /product= "transposase homologue"
FT                   /note= "homologous to the Tnp gene"
FT   CDS             complement (433880..434110)
FT                   /*tag= k
FT                   /standard_name= "ORF K11"
FT                   /product= "protein of unknown function"
FT                   /note= "homologous to the FixU gene"
FT   CDS             complement (434107..434433)
FT                   /*tag= l
FT                   /standard_name= "ORF K12"
FT                   /product= "protein of unknown function"
FT   CDS             complement (434517..434711)
FT                   /*tag= m
FT                   /standard_name= "ORF K13"
FT                   /product= "ferrodoxin/ferrodoxin-like protein"
FT                   /note= "homologous to the FdxN gene"
FT   CDS             complement (434753..436234)
FT                   /*tag= n
FT                   /standard_name= "ORF K14"
FT                   /gene= "nifB"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT   CDS             complement (436460..438130)
FT                   /*tag= o
FT                   /standard_name= "ORF K15"
FT                   /gene= "nifA"
FT                   /product= "positive regulator of nif, fix and other
FT                            genes"
FT   CDS             complement (438297..438590)
FT                   /*tag= p
FT                   /standard_name= "ORF K16"
FT                   /gene= "fixX"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (438605..439912)
FT                   /*tag= q
FT                   /standard_name= "ORF K17"
FT                   /gene= "fixC"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (439923..441032)
FT                   /*tag= r
FT                   /standard_name= "ORF K18"
FT                   /gene= "fixB"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (441042..441899)
FT                   /*tag= s
FT                   /standard_name= "ORF K19"
FT                   /gene= "fixA"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (442316..442636)
FT                   /*tag= t
FT                   /standard_name= "ORF K20
FT                   /product= "protein of unknown function"
FT   CDS             complement (443313..443879)
FT                   /*tag= u
FT                   /standard_name= "ORF K21"
FT                   /product= "protein of unknown function"
FT   CDS             444337..445029
FT                   /*tag= v
FT                   /standard_name= "ORF K22"
FT                   /product= "ferrodoxin-like protein"
FT                   /note= "homologous to the NifQ gene"
FT   CDS             445088..446602
FT                   /*tag= w
FT                   /standard_name= "ORF K23"
FT                   /gene= "dctA"
FT                   /product= "C4-dicarboxylate transport protein"
FT                   /note= "homologous to the DctAI gene"
FT   CDS             446599..447843
FT                   /*tag= x
FT                   /standard_name= "ORF L1"
FT                   /product= "cytochrome P450-like protein"
FT                   /note= "homologous to the CamC gene"
FT   CDS             447844..448500
FT                   /*tag= y
FT                   /standard_name= "ORF L2"
FT                   /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT                            protein"
FT                   /note= "homologous to the LinA gene"
FT   CDS             448497..450203
FT                   /*tag= z
FT                   /standard_name= "ORF L3"
FT                   /product= "putative protein with degradative function"
FT   CDS             450341..451396
FT                   /*tag= aa
FT                   /standard_name= "ORF L4"
FT                   /product= "luciferase alpha-subunit-like protein"
FT                   /note= "homologous to the LuxA gene"
FT   CDS             452980..454494
FT                   /*tag= ab
```

```
FT                   /standard_name= "ORF L6"
FT                   /gene= "nifD"
FT                   /product= "alpha-subunit of FeMo protein of nitrogenase"
FT   CDS             454590..456131
FT                   /*tag= ac
FT                   /standard_name= "ORF L7"
FT                   /gene= "nifK"
FT                   /product= "beta-subunit of FeMo protein of nitrogenase"
FT   CDS             456187..457677
FT                   /*tag= ad
FT                   /standard_name= "ORF L8"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT                   /note= "homologous to the NifB gene"
FT   CDS             457687..459096
FT                   /*tag= ae
FT                   /standard_name= "ORF L9"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT                   /note= "homologous to the FixF gene"
FT   CDS             459093..459575
FT                   /*tag= af
FT                   /standard_name= "ORF L10"
FT                   /product= "protein of unknown function"
FT                   /note= "homologous to the NifX gene"
FT   CDS             459579..460067
FT                   /*tag= ag
FT                   /standard_name= "ORF L11"
FT                   /product= "protein of unknown function"
FT   CDS             460501..460920
FT                   /*tag= ah
FT                   /standard_name= "ORF L12"
FT                   /product= "protein similar to part of the Fe protein
FT                            of nitrogenase"
FT                   /note= "homologous to the NifH gene"
FT   CDS             461228..461545
FT                   /*tag= ai
FT                   /standard_name= "ORF L13"
FT                   /product= "protein of unknown function"
FT   CDS             463201..464739
FT                   /*tag= aj
FT                   /standard_name= "ORF L14"
FT                   /product= "peptidase-like protein"
FT                   /note= "homologous to the bI-MPP gene"
FT   CDS             464736..466079
FT                   /*tag= ak
FT                   /standard_name= "ORF L15"
FT                   /product= "processing protease-like protein"
FT                   /note= "homologous to the PP gene"
FT   CDS             466590..467021
```

```
  Query Match             56.0%;  Score 14;  DB 19;  Length 534720;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtca 17
           ||||||||||||||
Db 435493 tcaggtgcaggtca 435506
```

```
RESULT  30
AAV30459
ID   AAV30459 standard; DNA; 536165 BP.
XX
AC   AAV30459;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Rhizobium species symbiotic plasmid pNGR234.
XX
KW   Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW   degradation; metabolism; host range; nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
PN   WO9802560-A2.
XX
PD   22-JAN-1998.
XX
PF   10-JUL-1997;   97WO-IB00950.
XX
PR   20-MAY-1997;   97GB-0010395.
PR   12-JUL-1996;   96EP-0730001.
XX
PA   (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
PA   (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
XX
PI   Broughton WJ,  Freiberg CB,  Perret XP,  Rosenthal A;
XX
DR   WPI; 1998-110606/10.
XX
PT   New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT   develop products for modifying plant characteristics, e.g. nitrogen
PT   fixation, synthesis of compounds and stress response
XX
PS   Claim 1; Fig 3; 228pp; English.
XX
CC   This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC   Rhizobium sp. NGR234.  Open reading frames (ORF) derivable from the
CC   nucleotide sequence are claimed.  The nucleotide sequences or ORFs can
CC   be used e.g. in the transportation of compounds to and from an organism
CC   which is a host to at least one of the nucleotide sequences, ORFs or
CC   proteins, the degradation and/or metabolism of organic, inorganic,
CC   natural or xenobiotic substances in a host organism or the modification
CC   of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC   minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC   especially for nodulation efficiency on host plants.
XX
SQ   Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;
```

```
  Query Match             56.0%;  Score 14;  DB 19;  Length 536165;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtca 17
           ||||||||||||||
Db 435493 tcaggtgcaggtca 435506
```

```
RESULT  31
AAQ13647
ID   AAQ13647 standard; DNA; 18 BP.
XX
AC   AAQ13647;
XX
DT   28-NOV-1991  (first entry)
XX
DE   Exon B/intron junction of PTPase B1 gene.
XX
KW   PTPB1; protein phosphotyrosyl phosphatase B1;
KW   growth suppression activity; ANLL; myeloproliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..9
FT                   /*tag= a
FT                   /note= "last 3 codons of exon B"
FT   intron          10..18
FT                   /*tag= b
FT                   /note= "5' end of intron"
XX
PN   WO9113173-A.
XX
```

```
PD   05-SEP-1991.
XX
PF   01-MAR-1991;   91WO-0001432.
XX
PR   18-JAN-1991;   91US-0643041.
PR   02-MAR-1990;   90US-0487733.
XX
PA   (BIOT-) APPL BIOTECHN INC.
XX
PI   Bruskin AM,  Hill DE;
XX
DR   WPI; 1991-281485/38.
XX
PT   Vector contg. protein phospho:tyrosyl phosphatase gene fragment -
PT   for use in diagnosis of cancers, e.g. acute non-lymphocytic
PT   leukaemia
XX
PS   Disclosure; Fig 4B; 73pp; English.
XX
CC   The PTPase B1 genomic sequence comprises a number of exons and
CC   introns. Exon B encodes amino acids Met(235) to Gln(288) of the
CC   PTPase B1 protein (see AAR13907). This is the junction of the 3' end
CC   of exon B and the 5' end of the intron.
CC   See AAQ13643-Q13654.
XX
SQ   Sequence 18 BP; 3 A; 5 C; 5 G; 5 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 12;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gtgcaggtcagca 20
           |||||||||||||
Db       4 gtgcaggtcagca 16


RESULT  32
AAX85925/c
ID   AAX85925 standard; DNA; 123 BP.
XX
AC   AAX85925;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Oligonucleotide used to produce heavy chain variable region of Ig NEW.
XX
KW   Light chain variable region; interleukin-4; IL-4; antibody 3B9;
KW   chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;
KW   Immunoglobulin E-mediated allergic reaction; allergic rhinitis;
KW   conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
KW   rheumatoid arthritis; host-versus-graft disease; renal disease;
KW   allergy; Ig NEW; ss.
XX
OS   Synthetic.
XX
PN   US5928904-A.
XX
PD   27-JUL-1999.
XX
PF   07-JUN-1995;   95US-0483632.
XX
PR   07-JUN-1995;   95US-0483632.
PR   07-SEP-1993;   93US-0117366.
PR   14-OCT-1993;   93US-0136783.
PR   07-SEP-1994;   94WO-US10308.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Gross MS,  Holmes SD,  Sylvester DR;
XX
DR   WPI; 1999-429500/36.
XX
PT   New DNA molecules encoding recombinant antibodies useful for
PT   treating IL4-mediated conditions
XX
PS   Disclosure; Columns 61-63; 50pp; English.
XX
CC   The specification describes chimeric and humanised IL-4
CC   monoclonal antibodies. The antibodies of the invention are used in
CC   therapeutic and pharmaceutical compositions for treating IL-4 mediated
CC   and Immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
CC   conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
CC   rheumatoid arthritis, host-versus-graft disease and renal disease.
CC   They are also useful in the diagnosis of an allergy or condition
CC   associated with excess IL-4 production through the measurement e.g. by
CC   ELISA of circulating endogenous IL-4 levels in humans. Oligonucleotides
CC   AAX85925-28 were used to produce the heavy chain variable region of
CC   Ig NEW, in the course of the invention.
XX
SQ   Sequence 123 BP; 25 A; 37 C; 41 G; 20 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  33
AAQ60177/c
ID   AAQ60177 standard; DNA; 225 BP.
XX
AC   AAQ60177;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST02167.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 4; Page 293; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST02167 has a "poor" coding probability as evaluated using the
```

```
CC   coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
SQ   Sequence 225 BP; 67 A; 67 C; 49 G; 41 T; 1 other;


  Query Match             52.0%;  Score 13;  DB 14;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 tgcaggtcagcac 21
          |||||||||||||
Db     96 TGCAGGTCAGCAC 84


RESULT  34
AAC00049/c
ID   AAC00049 standard; cDNA; 276 BP.
XX
AC   AAC00049;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 47.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00043.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 47; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 276 BP; 55 A; 93 C; 69 G; 59 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 276;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db    143 GGTGCAGGTCAGC 131


RESULT  35
AAX85929/c
ID   AAX85929 standard; DNA; 363 BP.
XX
AC   AAX85929;
XX
DT   13-SEP-1999  (first entry)
XX
DE   DNA encoding the heavy chain variable region of Ig NEW.
XX
KW   Heavy chain variable region; interleukin-4; IL-4; antibody 3B9;
KW   chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;
KW   Immunoglobulin E-mediated allergic reaction; allergic rhinitis;
KW   conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
KW   rheumatoid arthritis; host-versus-graft disease; renal disease;
KW   allergy; Ig NEW; ss.
XX
OS   Synthetic.
XX
PN   US5928904-A.
XX
PD   27-JUL-1999.
XX
PF   07-JUN-1995;   95US-0483632.
XX
PR   07-JUN-1995;   95US-0483632.
PR   07-SEP-1993;   93US-0117366.
PR   14-OCT-1993;   93US-0136783.
PR   07-SEP-1994;   94WO-US10308.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Gross MS,  Holmes SD,  Sylvester DR;
XX
DR   WPI; 1999-429500/36.
DR   P-PSDB; AAY23780.
XX
PT   New DNA molecules encoding recombinant antibodies useful for
PT   treating IL4-mediated conditions
XX
PS   Disclosure; Columns 63-65; 50pp; English.
XX
CC   The specification describes chimeric and humanised IL-4
CC   monoclonal antibodies. The antibodies of the invention are used in
CC   therapeutic and pharmaceutical compositions for treating IL-4 mediated
CC   and Immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
CC   conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
CC   rheumatoid arthritis, host-versus-graft disease and renal disease.
CC   They are also useful in the diagnosis of an allergy or condition
CC   associated with excess IL-4 production through the measurement e.g. by
CC   ELISA of circulating endogenous IL-4 levels in humans. The present
CC   sequence encodes the heavy chain variable region of Ig NEW, and is
CC   used in the course of the invention.
XX
SQ   Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  36
AAX79527/c
```

```
ID   AAX79527 standard; cDNA; 363 BP.
XX
AC   AAX79527;
XX
DT   11-AUG-1999  (first entry)
XX
DE   Heavy chain coding sequence for humanised 3B9 antibody.
XX
KW   Antibody; interleukin-4; IL4; immunoglobulin E; IgE mediated disease;
KW   allergic disorder; allergic rhinitis; conjunctivitis; atopic dermatitis;
KW   atopic asthma; anaphylactic shock; cell proliferation regulator; therapy;
KW   autoimmune disease; graft versus host disease; ss.
XX
OS   Synthetic.
XX
PN   US5914110-A.
XX
PD   22-JUN-1999.
XX
PF   07-JUN-1995;   95US-0483636.
XX
PR   07-JUN-1995;   95US-0483636.
PR   07-SEP-1993;   93US-0117366.
PR   14-OCT-1993;   93US-0136783.
PR   07-SEP-1994;   94WO-US10308.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Gross MS,  Holmes SD,  Sylvester DR;
XX
DR   WPI; 1999-370482/31.
DR   P-PSDB; AAY18122.
XX
PT   Recombinant IL4 antibodies
XX
PS   Example 3; Column 63-64; 50pp; English.
XX
CC   This sequence encodes the heavy chain of the humanised 3B9
CC   antibody of the invention. The antibody is a chimeric or
CC   humanised interleukin-4 (IL4) monoclonal antibody for the treatment of
CC   immunoglobulin E (IgE) mediated diseases. The antibodies are useful for
CC   the treatment of allergic disorders such as allergic rhinitis,
CC   conjunctivitis, atopic dermatitis, atopic asthma and anaphylactic shock.
CC   The antibodies are also useful for regulating B and T cell proliferation
CC   and as such are useful in the treatment of autoimmune diseases and graft
CC   versus host disease.
XX
SQ   Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  37
AAF22004
ID   AAF22004 standard; DNA; 375 BP.
XX
AC   AAF22004;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 391.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB59101.
XX
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 823; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 375 BP; 79 A; 77 C; 129 G; 71 T; 19 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 375;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggt 15
           |||||||||||||
Db     171 gtcaggtgcaggt 183


RESULT  38
AAT61237/c
ID   AAT61237 standard; DNA; 378 BP.
XX
AC   AAT61237;
XX
DT   13-MAY-1997  (first entry)
XX
DE   Human anti-RSV monoclonal antibody RF-2 VH domain DNA.
XX
KW   Monoclonal antibody; MAb; RF-1; RF-2; respiratory syncytial virus;
```

KW RSV; fusion protein; F-protein; vaccine; immunotherapy; therapy;
KW Epstein Barr virus; immortalisation; recombinant antibody; ss.
XX
OS Homo sapiens.
XX
PN WO9640252-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US10070.
XX
PR 07-JUN-1995; 95US-0488376.
XX
PA (IDEC-) IDEC PHARM CORP.
XX
PI Brams P, Chamat SS, Heard CJ, Newman RA, Pan L;
PI Walsh EE;
XX
DR WPI; 1997-099892/09.
DR P-PSDB; AAW11625.
XX
PT Human monoclonal antibody specific for respiratory syncytial virus
PT fusion protein - used for the prevention and treatment of RSV
PT infection
XX
PS Claim 9; Fig 7b; 85pp; English.
XX
CC Nucleic acids (AAT61236-39) respectively code for the RF-1 VL domain,
CC RF-2 VH domain, RF-2 VL domain and RF-1 VH domain (AAW11634-37) of
CC human monoclonal antibodies (hMAbs) RF-1 and RF-2, which show high
CC affinity for respiratory syncytial virus (RSV) fusion protein. The
CC hMAbs were obtd. by: antigen priming of naive human splenocytes in
CC vitro; transferral of the primed splenocytes to a SCID mouse;
CC boosting with antigen; isolating human antibody-producing B-cells
CC from the SCID mice; and immortalisation of the B-cells using
CC Epstein Barr virus. DNA encoding the RF-1 and RF-2 heavy and
CC light variable domains can be incorporated into vectors (e.g.
CC NEOSPLA) and used to transfect eukaryotic (e.g. CHO) cells (see
CC also AAT61240-42, AAT61279) to provide a constant, stable supply of
CC anti-RSV F-protein hMAbs for use in the treatment or prevention of
CC RSV infection.
CC (N.B. Fig 7b is referred to as the RF-1 VH domain in the
CC specification, but is identical to the RF-2 VH domain of Fig 11b).
XX
SQ Sequence 378 BP; 91 A; 114 C; 90 G; 83 T; 0 other;

Query Match 52.0%; Score 13; DB 18; Length 378;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcag 18
||||||||||||
Db 70 AGGTGCAGGTCAG 58

RESULT 39
AAT61239/c
ID AAT61239 standard; DNA; 378 BP.
XX
AC AAT61239;
XX
DT 13-MAY-1997 (first entry)
XX
DE Human anti-RSV monoclonal antibody RF-1 VH domain DNA.
XX
KW Monoclonal antibody; MAb; RF-1; RF-2; respiratory syncytial virus;
KW RSV; fusion protein; F-protein; vaccine; immunotherapy; therapy;
KW Epstein Barr virus; immortalisation; recombinant antibody; ss.
XX
OS Homo sapiens.
XX
PN WO9640252-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US10070.
XX
PR 07-JUN-1995; 95US-0488376.
XX
PA (IDEC-) IDEC PHARM CORP.
XX
PI Brams P, Chamat SS, Heard CJ, Newman RA, Pan L;
PI Walsh EE;
XX
DR WPI; 1997-099892/09.
XX
PT Human monoclonal antibody specific for respiratory syncytial virus
PT fusion protein - used for the prevention and treatment of RSV
PT infection
XX
PS Claim 9; Fig 8b; 85pp; English.
XX
CC Nucleic acids (AAT61236-39) respectively code for the RF-1 VL domain,
CC RF-2 VH domain, RF-2 VL domain and RF-1 VH domain (AAW11634-37) of
CC human monoclonal antibodies (hMAbs) RF-1 and RF-2, which show high
CC affinity for respiratory syncytial virus (RSV) fusion protein. The
CC hMAbs were obtd. by: antigen priming of naive human splenocytes in
CC vitro; transferral of the primed splenocytes to a SCID mouse;
CC boosting with antigen; isolating human antibody-producing B-cells
CC from the SCID mice; and immortalisation of the B-cells using
CC Epstein Barr virus. DNA encoding the RF-1 and RF-2 heavy and
CC light variable domains can be incorporated into vectors (e.g.
CC NEOSPLA) and used to transfect eukaryotic (e.g. CHO) cells (see
CC also AAT61240-42, AAT61279) to provide a constant, stable supply of
CC anti-RSV F-protein hMAbs for use in the treatment or prevention of
CC RSV infection.
CC (N.B. Fig 8b is referred to as the RF-2 VH domain in the
CC specification, but is identical to the RF-1 VH domain of Fig 9b).
XX
SQ Sequence 378 BP; 86 A; 110 C; 97 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 18; Length 378;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ggtgcaggtcagc 19
|||||||||||||
Db 69 GGTGCAGGTCAGC 57

RESULT 40
AAC00033/c
ID AAC00033 standard; cDNA; 400 BP.
XX
AC AAC00033;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 31.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.

```
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00027.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 31; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 400 BP; 90 A; 117 C; 89 G; 92 T; 12 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 400;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    146 AGGTGCAGGTCAG 134


RESULT  41
AAX79207/c
ID   AAX79207 standard; cDNA; 423 BP.
XX
AC   AAX79207;
XX
DT   17-AUG-1999  (first entry)
XX
DE   Anti-HIV-1 gp120 antibody 694-D VH chain gene.
XX
KW   Heavy chain; variable region; human; HIV-1; gp120; monoclonal antibody;
KW   epitope; V3 loop; heterohybridoma; human immunodeficiency virus-1;
KW   peripheral blood lymphocyte; Epstein-Barr virus; EBV; AIDS; ds.
XX
OS   Homo sapiens.
XX
PN   US5914109-A.
XX
PD   22-JUN-1999.
XX
PF   21-NOV-1994;   94US-0345321.
XX
PR   23-APR-1992;   92US-0872675.
PR   15-JUN-1990;   90US-0538451.
PR   12-APR-1991;   91US-0684090.
PR   21-NOV-1994;   94US-0345321.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Gorny MK,  Zolla-Pazner S;
XX
DR   WPI; 1999-370481/31.
DR   P-PSDB; AAY07486.
XX
PT   Heterohybridoma producing human monoclonal antibodies to human
PT   immunodeficiency virus-1
XX
PS   Example 8; Fig 13; 42pp; English.
XX
CC   This sequence represents the coding region for the heavy chain variable
CC   region of the human anti-HIV-1 gp120 monoclonal antibody 694-D.  The
CC   antibody is targeted to an epitope on the V3 loop of gp120.  The
CC   invention relates to the generation of heterohybridomas producing human
CC   monoclonal antibodies to a neutralising epitope of human immunodeficiency
CC   virus-1 (HIV-1) prepared by transforming peripheral blood lymphocytes
CC   with Epstein-Barr virus.  The antibodies can be used to treat someone
CC   infected with HIV-1 or suffering from AIDS.
XX
SQ   Sequence 423 BP; 96 A; 127 C; 104 G; 95 T; 1 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    127 AGGTGCAGGTCAG 115


RESULT  42
AAZ42289/c
ID   AAZ42289 standard; cDNA; 437 BP.
XX
AC   AAZ42289;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human 5' EST isolated from a cDNA library SEQ ID NO:48.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW   gene therapy; chromosome mapping; upstream regulatory sequence;
KW   forensic; location; development; protein synthesis; stability;
KW   regulation; identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO9953051-A2.
XX
PD   21-OCT-1999.
XX
PF   09-APR-1999;   99WO-IB00712.
XX
PR   09-APR-1998;   98US-0057719.
PR   28-APR-1998;   98US-0069047.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-038446/03.
DR   P-PSDB; AAY64675.
XX
PT   Novel secreted protein 5' expressed sequence tag sequences used in
PT   diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS   Claim 1; Page 198; 837pp; English.
XX
CC   AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC   sequences, corresponding to human secreted proteins. AAY64651 to
CC   AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC   AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC   products. They can be used to identify and isolate 5' untranslated
CC   regions (UTRs) and upstream regulatory regions which control the
CC   location, development stage, rate, and quantity of protein synthesis, as
CC   well as stability of mRNA. The ESTs are also useful as probes for
CC   chromosome mapping, and to obtain full length cDNA clones. The ESTs can
```

CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 437 BP; 94 A; 132 C; 105 G; 102 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 437;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcag 18
||||||||||||||
Db 146 AGGTGCAGGTCAG 134

RESULT 43
AAC75419
ID AAC75419 standard; cDNA; 472 BP.
XX
AC AAC75419;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF974 polynucleotide sequence SEQ ID NO:1947.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41210.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 1479-1480; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 472 BP; 94 A; 151 C; 174 G; 53 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 472;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgcaggtca 17
|||||||||||||
Db 185 caggtgcaggtca 197

RESULT 44
AAZ20408/c
ID AAZ20408 standard; cDNA; 477 BP.
XX
AC AAZ20408;
XX
DT 19-NOV-1999 (first entry)
XX
DE IgG antibody 2.3.2 heavy chain coding sequence.
XX
KW Antibody; CD147; IgM; ABX-CBL; activated T-cell killing; leukaemia;
KW activated B-cell; monocyte; graft versus host disease; therapy; cancer;
KW organ transplant rejection disease; lymphoma; pancreatic disease;
KW autoimmune disease; inflammatory disease; arthritis; binding site; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..447
FT /*tag= a
FT /note= "no stop codon given"
XX
PN WO9945031-A2.
XX
PD 10-SEP-1999.
XX
PF 03-MAR-1999; 99WO-US04583.
XX
PR 03-MAR-1998; 98US-0034607.
PR 03-FEB-1999; 99US-0244253.
XX
PA (ABGE-) ABGENIX INC.
XX
PI Davis CG, Blacher RW, Corvalan JR, Culwell AR, Green LL, Hales J;
PI Havrilla N, Ivanov VE, Lipani JA, Liu Q, Weber RF, Yang X;
XX
DR WPI; 1999-540816/45.

DR P-PSDB; AAY34307.
XX
PT New monoclonal antibody, used for treating e.g. graft versus host
PT disease, cancers, autoimmune diseases and inflammatory diseases -
XX
PS Disclosure; Fig 30; 245pp; English.
XX
CC This sequence encodes the heavy chain of an antibody of the
CC invention. The antibody is a monoclonal antibody (MAb) with an isotype
CC that fixes complement and a variable region that binds to the epitope on
CC CD147 bound by the IgM MAb ABX-CBL, providing that the antibody is not
CC CBL1. The MAb can selectively kill activated T-cells, activated B-cells
CC or resting or activated monocytes. The products and methods can be used
CC for treating diseases involving activated T-cells or B-cells or
CC monocytes, e.g. graft versus host disease (GVHD), organ transplant
CC rejection diseases (e.g. renal transplant, ocular transplant), cancers
CC (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
CC pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
CC (e.g. arthritis).
XX
SQ Sequence 477 BP; 97 A; 157 C; 119 G; 104 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 477;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcag 18
||||||||||||||
Db 13 AGGTGCAGGTCAG 1

RESULT 45
AAV87894
ID AAV87894 standard; cDNA; 527 BP.
XX
AC AAV87894;
XX
DT 12-FEB-1999 (first entry)
XX
DE EST clone FG380.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845437-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06956.
XX
PR 10-APR-1997; 97US-0837312.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070078/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
PS Claim 1; Page 215; 641pp; English.
XX
CC The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 527 BP; 108 A; 162 C; 139 G; 118 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 527;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ggtgcaggtcagc 19
||||||||||||||
Db 347 ggtgcaggtcagc 359

Search completed: October 9, 2001, 15:53:51
Job time: 15192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:42 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12575

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c	5	15	60.0	630	3	US-08-545-809A-10	Sequence 10, Appl
c	6	14	56.0	613	3	US-08-545-809A-5	Sequence 5, Appli
c	7	13	52.0	123	2	US-08-483-636-59	Sequence 59, Appl
c	8	13	52.0	123	2	US-08-483-632-59	Sequence 59, Appl
c	9	13	52.0	363	2	US-08-483-636-63	Sequence 63, Appl
c	10	13	52.0	363	2	US-08-483-632-63	Sequence 63, Appl
c	11	13	52.0	378	1	US-08-488-376-13	Sequence 13, Appl
c	12	13	52.0	378	1	US-08-488-376-15	Sequence 15, Appl
c	13	13	52.0	378	2	US-08-634-223-13	Sequence 13, Appl
c	14	13	52.0	378	2	US-08-634-223-15	Sequence 15, Appl
c	15	13	52.0	378	2	US-08-634-224-13	Sequence 13, Appl
c	16	13	52.0	378	2	US-08-634-224-15	Sequence 15, Appl
c	17	13	52.0	378	2	US-08-634-400-13	Sequence 13, Appl
c	18	13	52.0	378	2	US-08-634-400-15	Sequence 15, Appl
c	19	13	52.0	378	2	US-08-635-878-13	Sequence 13, Appl
c	20	13	52.0	378	2	US-08-635-878-15	Sequence 15, Appl
c	21	13	52.0	378	2	US-08-770-057-13	Sequence 13, Appl
c	22	13	52.0	378	2	US-08-770-057-15	Sequence 15, Appl
c	23	13	52.0	378	4	US-09-335-697B-13	Sequence 13, Appl
c	24	13	52.0	378	4	US-09-335-697B-15	Sequence 15, Appl
c	25	13	52.0	423	2	US-08-345-321-9	Sequence 9, Appli
c	26	13	52.0	546	3	US-08-545-809A-26	Sequence 26, Appl
	27	13	52.0	718	3	US-09-154-083-31	Sequence 31, Appl
c	28	13	52.0	879	3	US-08-714-071-3	Sequence 3, Appli
c	29	13	52.0	1173	3	US-08-706-216-5	Sequence 5, Appli
c	30	13	52.0	1428	1	US-08-488-376-17	Sequence 17, Appl
c	31	13	52.0	1428	1	US-08-488-376-19	Sequence 19, Appl
c	32	13	52.0	1428	2	US-08-634-223-17	Sequence 17, Appl
c	33	13	52.0	1428	2	US-08-634-223-19	Sequence 19, Appl
c	34	13	52.0	1428	2	US-08-634-224-17	Sequence 17, Appl
c	35	13	52.0	1428	2	US-08-634-224-19	Sequence 19, Appl
c	36	13	52.0	1428	2	US-08-634-400-17	Sequence 17, Appl
c	37	13	52.0	1428	2	US-08-634-400-19	Sequence 19, Appl
c	38	13	52.0	1428	2	US-08-635-878-17	Sequence 17, Appl
c	39	13	52.0	1428	2	US-08-635-878-19	Sequence 19, Appl
c	40	13	52.0	1428	2	US-08-770-057-17	Sequence 17, Appl
c	41	13	52.0	1428	2	US-08-770-057-19	Sequence 19, Appl
c	42	13	52.0	1428	4	US-09-335-697B-17	Sequence 17, Appl
c	43	13	52.0	1428	4	US-09-335-697B-19	Sequence 19, Appl
	44	13	52.0	2830	2	US-09-010-928B-1	Sequence 1, Appli
	45	13	52.0	3105	4	US-08-542-635-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7
```

```
  Query Match           100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match           100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;       LOCATION:  5098..5574
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /codon_start= 5098
;       OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;       OTHER INFORMATION:  /product= "protein"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /gene= "ORFI"
;       OTHER INFORMATION:  /number= 6
;     FEATURE:
;       NAME/KEY:  -10_signal
;       LOCATION:  45..49
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "promoter ptac"
;     FEATURE:
;       NAME/KEY:  -35_signal
;       LOCATION:  23..28
;       OTHER INFORMATION:  /standard_name= "promoter ptac"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  105..119
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;       OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  2284..2297
;       OTHER INFORMATION:  /standard_name= "bioC RBS"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  3742..3752
;       OTHER INFORMATION:  /standard_name= "bioA RBS"
;     FEATURE:
;       NAME/KEY:  RBS
;       LOCATION:  5088..5100
;       OTHER INFORMATION:  /standard_name= "ORFI RBS"
;     FEATURE:
;       NAME/KEY:  terminator
;       LOCATION:  5583..5644
;       OTHER INFORMATION:  /standard_name= "rho-independent
;       OTHER INFORMATION:  transcriptional terminator"
;     FEATURE:
;       NAME/KEY:  stem_loop
;       LOCATION:  5583..5605
;     FEATURE:
;       NAME/KEY:  promoter
;       LOCATION:  1..96
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  /function= "promoter ptac"
;       OTHER INFORMATION:  /evidence= EXPERIMENTAL
;     PUBLICATION INFORMATION:
;       DOCUMENT NUMBER:  WO 87/01391 B1
;       FILING DATE:  26-AUG-1986
;       PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
```

```
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   5
US-08-545-809A-10/c
; Sequence 10, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,809A
;      FILING DATE:  27-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00603
;      FILING DATE:  10-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06501/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  630 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  human lymphoblast
;      CELL LINE:  CGM1
US-08-545-809A-10


  Query Match             60.0%;  Score 15;  DB 3;  Length 630;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagca 20
           |||||||||||||||
Db     280 AGGTGCAGGTCAGCA 266


RESULT   6
US-08-545-809A-5/c
; Sequence 5, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,809A
;      FILING DATE:  27-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00603
;      FILING DATE:  10-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06501/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  613 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  human lymphoblast
;      CELL LINE:  CGM1
US-08-545-809A-5


  Query Match             56.0%;  Score 14;  DB 3;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     283 AGGTGCAGGTCAGC 270


RESULT   7
US-08-483-636-59/c
; Sequence 59, Application US/08483636
; Patent No. 5914110
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
```

```
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,636
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  123 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-636-59


  Query Match             52.0%;  Score 13;  DB 2;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT   8
US-08-483-632-59/c
; Sequence 59, Application US/08483632
; Patent No. 5928904
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,632
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  123 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-632-59


  Query Match             52.0%;  Score 13;  DB 2;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT   9
US-08-483-636-63/c
; Sequence 63, Application US/08483636
; Patent No. 5914110
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,636
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-483-636-63


  Query Match           52.0%;  Score 13;  DB 2;  Length 363;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  10
US-08-483-632-63/c
; Sequence 63, Application US/08483632
; Patent No. 5928904
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,632
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-483-632-63


  Query Match           52.0%;  Score 13;  DB 2;  Length 363;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  11
US-08-488-376-13/c
; Sequence 13, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-488-376-13


  Query Match           52.0%;  Score 13;  DB 1;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  12
US-08-488-376-15/c
; Sequence 15, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-488-376-15


  Query Match           52.0%;  Score 13;  DB 1;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  13
US-08-634-223-13/c
; Sequence 13, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-223-13


  Query Match           52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  14
US-08-634-223-15/c
; Sequence 15, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
```

```
;       STREET:  P.O. Box 1404
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  United States
;       ZIP:  22313-1404
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/634,223
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  378 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..378
US-08-634-223-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  15
US-08-634-224-13/c
; Sequence 13, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;     APPLICANT:  BRAMS, Peter
;     APPLICANT:  CHAMAT, Soulaima Salim
;     APPLICANT:  PAN, Li-Zhen
;     APPLICANT:  WALSH, Edward E.
;     APPLICANT:  HEARD, Cheryl Janne
;     APPLICANT:  NEWMAN, Roland Anthony
;     TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;     TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;     TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Burns, Doane, Swecker & Mathis
;       STREET:  P.O. Box 1404
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  United States
;       ZIP:  22313-1404
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/634,224
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  378 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..378
US-08-634-224-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  16
US-08-634-224-15/c
; Sequence 15, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;     APPLICANT:  BRAMS, Peter
;     APPLICANT:  CHAMAT, Soulaima Salim
;     APPLICANT:  PAN, Li-Zhen
;     APPLICANT:  WALSH, Edward E.
;     APPLICANT:  HEARD, Cheryl Janne
;     APPLICANT:  NEWMAN, Roland Anthony
;     TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;     TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;     TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Burns, Doane, Swecker & Mathis
;       STREET:  P.O. Box 1404
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  United States
;       ZIP:  22313-1404
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/634,224
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
```

```
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  378 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..378
US-08-634-224-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  17
US-08-634-400-13/c
; Sequence 13, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-400-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db     69 GGTGCAGGTCAGC 57


RESULT  18
US-08-634-400-15/c
; Sequence 15, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-400-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  19
US-08-635-878-13/c
; Sequence 13, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-635-878-13


  Query Match           52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  20
US-08-635-878-15/c
; Sequence 15, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-635-878-15


  Query Match           52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  21
US-08-770-057-13/c
; Sequence 13, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
```

```
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-770-057-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db     69 GGTGCAGGTCAGC 57


RESULT  22
US-08-770-057-15/c
; Sequence 15, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-770-057-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  23
US-09-335-697B-13/c
; Sequence 13, Application US/09335697B
; Patent No. 6200804
;  GENERAL INFORMATION:
;       APPLICANT: BRAMS, Peter
;                  CHAMAT, Soulaima Salim
;                  PAN, Li-Zhen
;                  WALSH, Edward E.
;                  HEARD, Cheryl Janne
;                  NEWMAN, Roland Anthony
;       TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                           MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                           METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREO
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/335,697B
;            FILING DATE: 06-Jul-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/770,057
;            FILING DATE: <Unknown>
```

```
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Teskin, Robin L.
;              REGISTRATION NUMBER: 35,030
;              REFERENCE/DOCKET NUMBER: 012712-150
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 13:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 378 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..378
;         SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-335-697B-13


  Query Match             52.0%;  Score 13;  DB 4;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  24
US-09-335-697B-15/c
; Sequence 15, Application US/09335697B
; Patent No. 6200804
;   GENERAL INFORMATION:
;         APPLICANT: BRAMS, Peter
;                    CHAMAT, Soulaima Salim
;                    PAN, Li-Zhen
;                    WALSH, Edward E.
;                    HEARD, Cheryl Janne
;                    NEWMAN, Roland Anthony
;         TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                             MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                             METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;         NUMBER OF SEQUENCES: 19
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Burns, Doane, Swecker & Mathis
;              STREET: P.O. Box 1404
;              CITY: Alexandria
;              STATE: Virginia
;              COUNTRY: United States
;              ZIP: 22313-1404
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/335,697B
;              FILING DATE: 06-Jul-2000
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/770,057
;              FILING DATE: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Teskin, Robin L.
;              REGISTRATION NUMBER: 35,030
;              REFERENCE/DOCKET NUMBER: 012712-150
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (703) 836-6620
;              TELEFAX: (703) 836-2021
;    INFORMATION FOR SEQ ID NO: 15:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 378 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         FEATURE:
;              NAME/KEY:  CDS
;              LOCATION:  1..378
;         SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-335-697B-15


  Query Match             52.0%;  Score 13;  DB 4;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  25
US-08-345-321-9/c
; Sequence 9, Application US/08345321
; Patent No. 5914109
;  GENERAL INFORMATION:
;    APPLICANT:  ZOLLA-PAZNER, Susan
;    APPLICANT:  GORNY, Miroslav K.
;    TITLE OF INVENTION:  HETEROHYBRIDOMAS PRODUCING HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES TO HIV-1
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/345,321
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,675
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Browdy, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  ZOLLA-PAZNER1B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..423
US-08-345-321-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 81;
```

```
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    127 AGGTGCAGGTCAG 115


RESULT  26
US-08-545-809A-26/c
; Sequence 26, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,809A
;      FILING DATE:  27-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00603
;      FILING DATE:  10-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06501/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  546 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  human lymphoblast
;      CELL LINE:  CGM1
US-08-545-809A-26


  Query Match             52.0%;  Score 13;  DB 3;  Length 546;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db    232 GGTGCAGGTCAGC 220


RESULT  27
US-09-154-083-31
; Sequence 31, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
;  APPLICANT: Wu, Kai
;  TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
;  TITLE OF INVENTION: Constructs Therefor
;  FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
;  CURRENT APPLICATION NUMBER: US/09/154,083
;  CURRENT FILING DATE: 1998-09-16
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-09-154-083-31


  Query Match             52.0%;  Score 13;  DB 3;  Length 718;
  Best Local Similarity   100.0%;  Pred. No. 79;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 gcaggtcagcacg 22
          |||||||||||||
Db    347 gcaggtcagcacg 359


RESULT  28
US-08-714-071-3/c
; Sequence 3, Application US/08714071
; Patent No. 6136584
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Shiro, OKUNO
;    APPLICANT:  Hisanobu, HIRANO
;    APPLICANT:  Sadahito, SHIN
;    TITLE OF INVENTION:  FK506 BINDING PROTEIN GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,071
;      FILING DATE:
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  879 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Human fetal brain cDNA library
;    IMMEDIATE SOURCE:
;      CLONE:  OTK4(6-1)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..393
US-08-714-071-3


  Query Match             52.0%;  Score 13;  DB 3;  Length 879;
```

```
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     303 GGTGCAGGTCAGC 291


RESULT  29
US-08-706-216-5/c
; Sequence 5, Application US/08706216
; Patent No. 6140098
;  GENERAL INFORMATION:
;    APPLICANT:  Balasubramanian, Sriram
;    APPLICANT:  Ford, John
;    APPLICANT:  Gorman, Daniel M.
;    APPLICANT:  Zurawski, Gerard
;    TITLE OF INVENTION:  MAMMALIAN PROTEASES; RELATED REAGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,216
;      FILING DATE:  30-AUG-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0613
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1173 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1173
US-08-706-216-5


  Query Match             52.0%;  Score 13;  DB 3;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtc 16
           |||||||||||||
Db     418 TCAGGTGCAGGTC 406


RESULT  30
US-08-488-376-17/c
; Sequence 17, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-488-376-17


  Query Match             52.0%;  Score 13;  DB 1;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  31
US-08-488-376-19/c
; Sequence 19, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-488-376-19


  Query Match             52.0%;  Score 13;  DB 1;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    127 AGGTGCAGGTCAG 115


RESULT  32
US-08-634-223-17/c
; Sequence 17, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-223-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db    126 GGTGCAGGTCAGC 114


RESULT  33
US-08-634-223-19/c
; Sequence 19, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1428
US-08-634-223-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  34
US-08-634-224-17/c
; Sequence 17, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,224
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-224-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  35
US-08-634-224-19/c
; Sequence 19, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,224
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-224-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  36
US-08-634-400-17/c
```

```
; Sequence 17, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-400-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  37
US-08-634-400-19/c
; Sequence 19, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-400-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  38
US-08-635-878-17/c
; Sequence 17, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-635-878-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  39
US-08-635-878-19/c
; Sequence 19, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-635-878-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  40
US-08-770-057-17/c
; Sequence 17, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-770-057-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  41
US-08-770-057-19/c
; Sequence 19, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-770-057-19
```

```
  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  42
US-09-335-697B-17/c
; Sequence 17, Application US/09335697B
; Patent No. 6200804
;  GENERAL INFORMATION:
;       APPLICANT: BRAMS, Peter
;                  CHAMAT, Soulaima Salim
;                  PAN, Li-Zhen
;                  WALSH, Edward E.
;                  HEARD, Cheryl Janne
;                  NEWMAN, Roland Anthony
;       TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                           MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                           METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREO
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/335,697B
;            FILING DATE: 06-Jul-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/770,057
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Teskin, Robin L.
;            REGISTRATION NUMBER: 35,030
;            REFERENCE/DOCKET NUMBER: 012712-150
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1428 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1428
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-335-697B-17


  Query Match             52.0%;  Score 13;  DB 4;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114
```

```
RESULT  43
US-09-335-697B-19/c
; Sequence 19, Application US/09335697B
; Patent No. 6200804
;  GENERAL INFORMATION:
;       APPLICANT: BRAMS, Peter
;                  CHAMAT, Soulaima Salim
;                  PAN, Li-Zhen
;                  WALSH, Edward E.
;                  HEARD, Cheryl Janne
;                  NEWMAN, Roland Anthony
;       TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                           MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                           METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Burns, Doane, Swecker & Mathis
;            STREET: P.O. Box 1404
;            CITY: Alexandria
;            STATE: Virginia
;            COUNTRY: United States
;            ZIP: 22313-1404
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/335,697B
;            FILING DATE: 06-Jul-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/770,057
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Teskin, Robin L.
;            REGISTRATION NUMBER: 35,030
;            REFERENCE/DOCKET NUMBER: 012712-150
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (703) 836-6620
;            TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 19:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1428 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  1..1428
;       SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-335-697B-19


  Query Match           52.0%;  Score 13;  DB 4;  Length 1428;
  Best Local Similarity  100.0%;  Pred. No. 76;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    127 AGGTGCAGGTCAG 115


RESULT  44
US-09-010-928B-1
; Sequence 1, Application US/09010928B
; Patent No. 5994099
;  GENERAL INFORMATION:
;    APPLICANT: ,Lewis, Randolph V
;    APPLICANT:  Hayashi, Cheryl Y
;    TITLE OF INVENTION:  EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
;    TITLE OF INVENTION:  CODING THEREFOR
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  8110 GATEHOUSE RD. SUITE 500E
;      CITY:  FALLS CHURCH
;      STATE:  VIRGINIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/010,928B
;      FILING DATE:  22-JAN-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28977
;      REFERENCE/DOCKET NUMBER:  1447-109P
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2830 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..2830
;      OTHER INFORMATION:  /note= "Flagelliform DNA sequence
;      OTHER INFORMATION:  taken from the 5' region.  The putative start codon is at
;      OTHER INFORMATION:  position 219"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  219..2830
US-09-010-928B-1


  Query Match           52.0%;  Score 13;  DB 2;  Length 2830;
  Best Local Similarity  100.0%;  Pred. No. 72;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtc 16
          |||||||||||||
Db    588 TCAGGTGCAGGTC 600


RESULT  45
US-08-542-635-1
; Sequence 1, Application US/08542635
; Patent No. 6218356
;  GENERAL INFORMATION:
;    APPLICANT:  Pawson, Anthony
;    APPLICANT:  Henkemeyer, Mark
;    APPLICANT:  Letwin, Kenneth
;    TITLE OF INVENTION:  NOVEL NEURAL RECEPTOR
;    TITLE OF INVENTION:  TYROSINE KINASE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bereskin & Parr
;      STREET:  40 King Street West, Box 401
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/542,635
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McDiarmid, Shona S.
;       REGISTRATION NUMBER:  38,798
;       REFERENCE/DOCKET NUMBER:  3153-162
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (416) 364-7311
;       TELEFAX:  (416) 361-1398
;       TELEX:  06-23115
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3105 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Mus musculus
;       DEVELOPMENTAL STAGE:  Embryo
;     IMMEDIATE SOURCE:
;       LIBRARY:  lambda gt10 cDNA library
;       CLONE:  Combined PnUKRACE A2 and K2 AND cDNA clones
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:  Distal end of chromosome 4
;       MAP POSITION:  near the ahd-1 mutation
US-08-542-635-1


  Query Match             52.0%;  Score 13;  DB 4;  Length 3105;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcac 21
           |||||||||||||
Db    3061 TGCAGGTCAGCAC 3073


Search completed: October  9, 2001, 15:55:43
Job time: 13219 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:39 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 513911

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	17	68.0	324	149	BF464374	BF464374 UI-M-CG0p
c 2	17	68.0	413	141	BE862745	UI-M-BH0-
c 3	17	68.0	495	24	AI789420	AI789420 uk99a06.y
c 4	17	68.0	641	243	AZ426129	AZ426129 1M0206H03
c 5	16	64.0	163	1	AA065941	AA065941 ml51a01.r
c 6	16	64.0	288	160	BB516544	BB516544 BB516544
7	16	64.0	328	190	W54143	W54143 md13c03.r1
8	16	64.0	338	190	W34110	W34110 mc57c06.r1
c 9	16	64.0	345	31	AV633196	AV633196 AV633196
c 10	16	64.0	368	2	AA110497	AA110497 ml62e03.r
11	16	64.0	370	121	AW822106	AW822106 uq13c09.x
12	16	64.0	395	148	BF444068	BF444068 261798 MA
13	16	64.0	398	226	AQ298728	AQ298728 HS_3110_B
c 14	16	64.0	430	20	AI467556	AI467556 ve36g09.x
15	16	64.0	544	122	AW892602	AW892602 CM3-NN000
c 16	16	64.0	553	229	AQ530462	AQ530462 RPCI-11-3
c 17	16	64.0	775	136	BE535032	BE535032 601233227
c 18	16	64.0	822	77	BE194624	Be194624 HVSMEh008
19	16	64.0	897	220	CNS021CM	AL176719 Tetraodon
c 20	16	64.0	976	174	BG172968	BG172968 602336627
c 21	16	64.0	1011	192	AK010837	AK010837 Mus muscu
c 22	16	64.0	1298	152	BG328084	BG328084 602427134
c 23	15	60.0	180	10	AA699872	AA699872 zj81f04.s
c 24	15	60.0	264	157	H22139	H22139 yl38a03.s1
c 25	15	60.0	279	187	R50061	R50061 yj59c10.s1
c 26	15	60.0	292	19	AI331540	AI331540 fa94b01.y
c 27	15	60.0	293	251	AZ877721	AZ877721 RPCI-23-1
c 28	15	60.0	300	127	BB175346	BB175346 BB175346
c 29	15	60.0	306	173	BG100182	BG100182 ux94g04.y
30	15	60.0	382	16	AI138808	AI138808 qd98g02.x
c 31	15	60.0	385	117	AW522699	AW522699 UI-R-BO0-
c 32	15	60.0	396	116	AW486040	AW486040 69428 MAR
c 33	15	60.0	407	1	AA062258	AA062258 ml55g08.r
c 34	15	60.0	407	151	BF599674	BF599674 263546 MA
c 35	15	60.0	408	224	AQ085001	AQ085001 HS_2269_A
c 36	15	60.0	409	151	BF653430	BF653430 277257 MA
c 37	15	60.0	416	2	AA146146	AA146146 mq65d12.r
c 38	15	60.0	446	13	AA881231	AA881231 vx10d04.r
c 39	15	60.0	457	151	BF604067	BF604067 269755 MA
c 40	15	60.0	477	235	AQ921326	AQ921326 RPCI-23-2
c 41	15	60.0	492	140	BE808281	BE808281 213447 MA
c 42	15	60.0	495	240	AZ261270	AZ261270 RPCI-23-4
c 43	15	60.0	496	151	BF652819	BF652819 276504 MA
44	15	60.0	525	150	BF567735	BF567735 UI-R-BO0-
c 45	15	60.0	529	114	AW290780	AW290780 NXNV046H0

ALIGNMENTS

```
RESULT   1
BF464374/c
LOCUS       BF464374      324 bp    mRNA            EST       04-DEC-2000
DEFINITION  UI-M-CG0p-bog-f-09-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
            UI-M-CG0p-bog-f-09-0-UI 3', mRNA sequence.
ACCESSION   BF464374
VERSION     BF464374.1  GI:11533557
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            The following repetitive elements were found in this cDNA sequence:
            102-169, >(GAA)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CG0p-bog-f-09-0-UI"
                     /clone_lib="NIH_BMAP_Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived,  please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_SEQ=None found"
BASE COUNT       91 a     84 c     89 g     60 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 149;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
Db     324 CGTCAGGTGCAGGTCAG 308


RESULT   2
BE862745/c
LOCUS       BE862745      413 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-BH0-ajf-b-11-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-ajf-b-11-0-UI 5', mRNA sequence.
ACCESSION   BE862745
VERSION     BE862745.1  GI:10382015
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
```

```
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .413
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-BH0-ajf-b-11-0-UI"
                      /clone_lib="NIH_BMAP_M_S1"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_M_S1 library is a subtracted library derived from
                      a mixture of normalized libraries from ten regions of the
                      mouse brain (cerebellum, brain stems, olfactory bulbs,
                      hypothalamus, cortex, amygdala, basal ganglia, pineal
                      gland, striatum, hipoccampus). The driver used for
                      subtraction consisted of a pool of 20,000 cDNA clones
                      obtained from non-normalized and normalized libraries of
                      these ten regions of the mouse brain."
BASE COUNT       92 a    125 c    112 g     84 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 141;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
Db      17 CGTCAGGTGCAGGTCAG 1


RESULT   3
AI789420/c
LOCUS       AI789420      495 bp    mRNA            EST       02-JUL-1999
DEFINITION  uk99a06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2064658 5', mRNA sequence.
ACCESSION   AI789420
VERSION     AI789420.1  GI:5337136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uk99a06.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:993846
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES              Location/Qualifiers
     source           1. .495
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:2064658"
                      /clone_lib="Sugano mouse kidney mkia"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                      (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                      was primed with an oligo(dT) primer
                      [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                      ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                      and cloned into distinct DraIII sites of the pME18S-FL3
                      vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                      be used to isolate the cDNA insert.  Size selection was
                      performed to exclude fragments <1.5kb.  Library
                      constructed by Dr. Sumio Sugano (University of Tokyo
                      Institute of Medical Science).  Custom primers  for
                      sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                      primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      147 a    122 c    141 g     85 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 24;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
Db     393 CGTCAGGTGCAGGTCAG 377


RESULT   4
AZ426129/c
LOCUS       AZ426129      641 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0206H03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0206H03 R, DNA sequence.
ACCESSION   AZ426129
VERSION     AZ426129.1  GI:10550142
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0206  row: H  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 641.
FEATURES              Location/Qualifiers
     source           1. .641
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0206H03"
```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 156 a 184 c 183 g 118 t
ORIGIN

Query Match 68.0%; Score 17; DB 243; Length 641;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcag 18
||||||||||||||||||
Db 212 CGTCAGGTGCAGGTCAG 196

RESULT 5
AA065941/c
LOCUS AA065941 163 bp mRNA EST 03-FEB-1997
DEFINITION m151a01.r1 Stratagene mouse testis (#937308) Mus musculus cDNA clone IMAGE:515496 5', mRNA sequence.
ACCESSION AA065941
VERSION AA065941.1 GI:1562644
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 163)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:309344
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 144.
FEATURES Location/Qualifiers
source 1. .163
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:515496"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 23 a 56 c 49 g 35 t
ORIGIN

Query Match 64.0%; Score 16; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gtcaggtgcaggtcag 18
||||||||||||||||
Db 16 GTCAGGTGCAGGTCAG 1

RESULT 6
BB516544/c
LOCUS BB516544 288 bp mRNA EST 28-JUL-2000
DEFINITION BB516544 RIKEN full-length enriched, 16 days neonate heart Mus musculus cDNA clone D830016J11 3' similar to M76601 Mouse alpha cardiac myosin heavy chain mRNA, mRNA sequence.
ACCESSION BB516544
VERSION BB516544.1 GI:9568002
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 288)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

```
          19-44 (1999)
           Please visit our web site (http://genome.rtc.riken.go.jp) for
          further details.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D830016J11"
                     /clone_lib="RIKEN full-length enriched, 16 days neonate
                     heart"
                     /tissue_type="heart"
                     /dev_stage="16 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       83 a    104 c     57 g     44 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 160;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtca 17
           ||||||||||||||||
Db     188 CGTCAGGTGCAGGTCA 173


RESULT   7
W54143
LOCUS       W54143         328 bp    mRNA            EST       03-JUN-1996
DEFINITION  md13c03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:368260 5', mRNA sequence.
ACCESSION   W54143
VERSION     W54143.1  GI:1355214
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:229692
            Seq primer: mob.REGA+ET.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:368260"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       81 a     66 c     85 g     96 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 190;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagc 19
           ||||||||||||||||
Db      86 TCAGGTGCAGGTCAGC 101


RESULT   8
W34110
LOCUS       W34110         338 bp    mRNA            EST       13-MAY-1996
DEFINITION  mc57c06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:352618 5', mRNA sequence.
ACCESSION   W34110
VERSION     W34110.1  GI:1316081
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:224418
            Putative full length read
            Seq primer: mob.REGA+ET
            High quality sequence stop: 324.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
```

```
                     /clone="IMAGE:352618"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       86 a     65 c     84 g    103 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 190;  Length 338;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagc 19
          ||||||||||||||||
Db     82 TCAGGTGCAGGTCAGC 97


RESULT  9
AV633196/c
LOCUS       AV633196      345 bp    mRNA            EST       15-DEC-2000
DEFINITION  AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HC017h07_r 5', mRNA sequence.
ACCESSION   AV633196
VERSION     AV633196.1  GI:10776516
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HC017h07_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       62 a    105 c     98 g     79 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 31;  Length 345;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db    173 CAGGTGCAGGTCAGCA 158


RESULT  10
AA110497/c
LOCUS       AA110497      368 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml62e03.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:516604 5', mRNA sequence.
ACCESSION   AA110497
VERSION     AA110497.1  GI:1662274
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:310452
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:516604"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       75 a    115 c    106 g     72 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 368;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     21 GTCAGGTGCAGGTCAG 6


RESULT  11
AW822106
LOCUS       AW822106      370 bp    mRNA            EST       17-MAY-2000
DEFINITION  uq13c09.x1 Ren Stubbs mouse thymus Mus musculus cDNA clone
            IMAGE:2802352 3', mRNA sequence.
ACCESSION   AW822106
VERSION     AW822106.1  GI:7915183
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uq13c09.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1041876
            Seq primer: Primer name ambiguous
            High quality sequence stop: 303.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2802352"
                     /clone_lib="Ren Stubbs mouse thymus"
                     /sex="mixed"
                     /dev_stage="3 weeks"
                     /lab_host="DH10B"
                     /note="Organ: thymus; Vector: pT7T3D-Pac; Site_1: NotI;
                     Site_2: PacI; 1st strand cDNA was primed with an oligo(dT)
                     primer; double-stranded cDNA was ligated using 5' linker
                     ggccgctat and 3' linker aactggaagcttaatt. Library is
                     size-selected >2.5 kb and average insert size is 3.5 kb.
                     Clones were arrayed from primary plating; non-amplified.
                     Library constructed by X. Ren and L. Stubbs (Lawrence
                     Livermore National Laboratory and DOE Joint Genome
                     Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
BASE COUNT      125 a     57 c     75 g    113 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 121;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db      53 GTCAGGTGCAGGTCAG 68


RESULT  12
BF444068
LOCUS       BF444068      395 bp    mRNA            EST       01-DEC-2000
DEFINITION  261798 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF444068
VERSION     BF444068.1  GI:11504160
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 96  row: D  column: 1
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      118 a     66 c    143 g     68 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 148;  Length 395;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db     309 CAGGTGCAGGTCAGCA 324


RESULT  13
AQ298728
LOCUS       AQ298728      398 bp    DNA             GSS       15-DEC-1998
DEFINITION  HS_3110_B1_C12_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3110 Col=23 Row=F, DNA sequence.
ACCESSION   AQ298728
VERSION     AQ298728.1  GI:4015907
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3110  row: F  column: 23
            Class: BAC ends
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="Plate=3110 Col=23 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       87 a     93 c    102 g    116 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 226;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacg 22
          ||||||||||||||||
Db    330 GGTGCAGGTCAGCACG 345


RESULT  14
AI467556/c
LOCUS       AI467556      430 bp    mRNA            EST       09-MAR-1999
DEFINITION  ve36g09.x1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:820288 3', mRNA sequence.
ACCESSION   AI467556
VERSION     AI467556.1  GI:4320893
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:488568
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:820288"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      136 a    102 c     85 g    107 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 20;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagc 19
          ||||||||||||||||
Db    259 TCAGGTGCAGGTCAGC 244


RESULT  15
AW892602
LOCUS       AW892602      544 bp    mRNA            EST       24-MAY-2000
DEFINITION  CM3-NN0004-100300-111-f07 NN0004 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW892602
VERSION     AW892602.1  GI:8056807
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-NN0004-100
            300-111-f07&t3=2000-03-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN0004"
                     /dev_stage="Adult"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      127 a    152 c    160 g    105 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 122;  Length 544;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db    383 GTCAGGTGCAGGTCAG 398


RESULT  16
AQ530462/c
LOCUS       AQ530462      553 bp    DNA             GSS       18-MAY-1999
DEFINITION  RPCI-11-369F11.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-369F11
            , DNA sequence.
```

```
ACCESSION   AQ530462
VERSION     AQ530462.1  GI:4842505
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-369F11.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends. .
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="GDB:7641442"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-369F11"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      108 a    165 c    157 g    123 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 229;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     204 GTCAGGTGCAGGTCAG 189


RESULT  17
BE535032/c
LOCUS       BE535032      775 bp    mRNA            EST       09-AUG-2000
DEFINITION  601233227F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3596935 5',
            mRNA sequence.
ACCESSION   BE535032
VERSION     BE535032.1  GI:9763677
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8775  row: f  column: 08
            High quality sequence stop: 225.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3596935"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      170 a    186 c    221 g    198 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 136;  Length 775;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtca 17
           ||||||||||||||||
Db     201 CGTCAGGTGCAGGTCA 186


RESULT  18
BE194624/c
ID   BE194624   standard; RNA; EST; 822 BP.
XX
AC   BE194624;
XX
SV   BE194624.1
XX
DT   05-JUL-2000 (Rel. 64, Created)
DT   30-JUL-2000 (Rel. 64, Last updated, Version 2)
XX
DE   HVSMEh0086E11f Hordeum vulgare 5-45 DAP spike EST library HVcDNA0009 (5' to
DE   45 DAP) Hordeum vulgare cDNA clone HVSMEh0086E11f, mRNA sequence.
XX
KW   EST.
XX
OS   Hordeum vulgare (barley)
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
XX
RN   [1]
RP   1-822
RA   Wing R., Close T.J., Kleinhofs A., Wise R., Begum D., Frisch D., Yu  Y.,
RA   Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo  T.,
RA   Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;
RT   "Development of a genetically and physically anchored EST resource for
RT   barley genomics";
RL   Unpublished.
XX
CC   Contact: Wing RA
CC   Clemson University Genomics Institute
CC   Clemson University
CC   100 Jordan Hall, Clemson, SC 29634, USA
CC   Tel: 864 656 7288
CC   Fax: 864 656 4293
CC   Email: rwing@clemson.edu
CC   Seq primer: AATTAACCCTCACTAAAGGG
CC   High quality sequence start: 225
```

```
CC   High quality sequence stop: 345.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .822
FT                   /db_xref="taxon:4513"
FT                   /db_xref="ESTLIB:5341"
FT                   /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1"
FT                   /organism="Hordeum vulgare"
FT                   /cultivar="Morex"
FT                   /clone="HVSMEh0086E11f"
FT                   /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
FT                   HVcDNA0009 (5 to 45 DAP)"
FT                   /tissue_type="5-45 DAP Spike"
FT                   /lab_host="SOLR"
XX
SQ   Sequence 822 BP; 276 A; 160 C; 184 G; 200 T; 2 other;


  Query Match             64.0%;  Score 16;  DB 77;  Length 822;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db    362 GTCAGGTGCAGGTCAG 347


RESULT  19
CNS021CM
LOCUS       CNS021CM      897 bp    DNA              GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            225C12 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL176719
VERSION     AL176719.1  GI:7814776
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 897)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="225C12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG225BB06LP1~end : T7"
BASE COUNT      118 a    284 c    294 g    182 t     19 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 220;  Length 897;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     16 GTCAGGTGCAGGTCAG 31


RESULT  20
BG172968/c
LOCUS       BG172968      976 bp    mRNA             EST       06-FEB-2001
DEFINITION  602336627F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459758 5',
            mRNA sequence.
ACCESSION   BG172968
VERSION     BG172968.1  GI:12679671
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D..
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10260  row: e  column: 07
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4459758"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      255 a    250 c    163 g    308 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 174;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 69;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db    168 GTCAGGTGCAGGTCAG 153


RESULT  21
AK010837/c
LOCUS       AK010837      1011 bp    mRNA             HTC       08-FEB-2001
DEFINITION  Mus musculus ES cells cDNA, RIKEN full-length enriched library,
            clone:2410198H06, full insert sequence.
ACCESSION   AK010837
VERSION     AK010837.1  GI:12846558
```

```
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2410198H06.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods Enzymol. 303, 19-44 (1999)
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Kikuchi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T.,
            Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T.,
            Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Okazaki,Y., Muramatsu,M., Inoue,Y. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1011)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1011
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1910427"
                     /db_xref="MGD:MGI:1924031"
                     /clone="2410198H06"
                     /cell_type="ES cells"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
BASE COUNT      255 a    227 c    180 g    349 t
ORIGIN
```

```
  Query Match           64.0%;  Score 16;  DB 192;  Length 1011;
  Best Local Similarity 100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     290 GTCAGGTGCAGGTCAG 275


RESULT  22
BG328084/c
LOCUS       BG328084     1298 bp    mRNA            EST       27-FEB-2001
DEFINITION  602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
            mRNA sequence.
ACCESSION   BG328084
VERSION     BG328084.1  GI:13134522
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1298)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1233  row: c  column: 10
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .1298
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4546497"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      316 a    498 c    357 g    127 t
ORIGIN
```

```
  Query Match           64.0%;  Score 16;  DB 152;  Length 1298;
  Best Local Similarity 100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtc 16
           ||||||||||||||||
Db     984 TCGTCAGGTGCAGGTC 969


RESULT  23
AA699872/c
LOCUS       AA699872     180 bp    mRNA            EST       19-DEC-1997
DEFINITION  zj81f04.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:461311 3', mRNA sequence.
ACCESSION   AA699872
VERSION     AA699872.1  GI:2702835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 144.
FEATURES             Location/Qualifiers
     source          1. .180
                     /organism="Homo sapiens"
                     /db_xref="GDB:3752334"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:461311"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       47 a     57 c     36 g     40 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 180;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagc 19
          |||||||||||||||
Db    138 CAGGTGCAGGTCAGC 124


RESULT  24
H22139/c
LOCUS       H22139        264 bp    mRNA            EST       06-JUL-1995
DEFINITION  yl38a03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone
            IMAGE:160492 3', mRNA sequence.
ACCESSION   H22139
VERSION     H22139.1  GI:890834
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 685
            High quality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 685   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 241.
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Homo sapiens"
                     /db_xref="GDB:574535"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:160492"
                     /clone_lib="Soares breast 3NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     20. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       87 a     47 c     50 g     73 t      7 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 157;  Length 264;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcag 18
          |||||||||||||||
Db    215 TCAGGTGCAGGTCAG 201


RESULT  25
R50061/c
LOCUS       R50061        279 bp    mRNA            EST       18-MAY-1995
DEFINITION  yj59c10.s1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:153042 3', mRNA sequence.
ACCESSION   R50061
VERSION     R50061.1  GI:811963
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 681
            High quality sequence stops: 257 Source: IMAGE Consortium, LLNL
```

```
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 681   Std Error: 0.00
            Seq primer: Promega -21ml3
            High quality sequence stop: 257.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Homo sapiens"
                     /db_xref="GDB:565317"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:153042"
                     /clone_lib="Soares breast 2NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     230. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       93 a     49 c     54 g     75 t      8 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 187;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcag 18
           ||||||||||||||||
Db     229 TCAGGTGCAGGTCAG 215


RESULT  26
AI331540/c
LOCUS       AI331540      292 bp    mRNA            EST       28-DEC-1998
DEFINITION  fa94b01.y1 zebrafish fin day1 regeneration Danio rerio cDNA 5',
            mRNA sequence.
ACCESSION   AI331540
VERSION     AI331540.1  GI:4068099
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="zebrafish fin day1 regeneration"
                     /sex="mixed male and female"
                     /tissue_type="1 day fin regenerates"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; 1st
                     strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed
                     by second strand synthesis, and ligated to 5' adapter (5'
                     )-aattcggcacgag- 3', 3'-gccgtgctc-5'.  cDNA was cloned
                     directionally (EcoRI/XhoI) into Stratagene Zap express
                     lambda phage arms.  Mass invivo excision done to obtain
                     inserts in pBK-CMV phagemid."
BASE COUNT       55 a     83 c     96 g     58 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 19;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcac 21
           ||||||||||||||||
Db     211 GGTGCAGGTCAGCAC 197


RESULT  27
AZ877721/c
LOCUS       AZ877721      293 bp    DNA             GSS       05-MAR-2001
DEFINITION  RPCI-23-195I20.TV RPCI-23 Mus musculus genomic clone RPCI-23-195I20
            , DNA sequence.
ACCESSION   AZ877721
VERSION     AZ877721.1  GI:13195787
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-195I20.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 195  row: I  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-195I20"
                     /clone_lib="RPCI-23"
                     /sex="Female"
```

/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 61 a 74 c 78 g 80 t
ORIGIN

Query Match 60.0%; Score 15; DB 251; Length 293;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gtcaggtgcaggtca 17
||||||||||||||||
Db 55 GTCAGGTGCAGGTCA 41

RESULT 28
BB175346/c
LOCUS BB175346 300 bp mRNA EST 29-JUN-2000
DEFINITION BB175346 RIKEN full-length enriched, adult male hypothalamus Mus musculus cDNA clone A230056E21 3', mRNA sequence.
ACCESSION BB175346
VERSION BB175346.1 GI:8834429
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 300)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .300
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A230056E21"
/clone_lib="RIKEN full-length enriched, adult male hypothalamus"
/sex="male"
/tissue_type="hypothalamus"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. "
BASE COUNT 66 a 81 c 58 g 95 t
ORIGIN

Query Match 60.0%; Score 15; DB 127; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcaggtgcaggtcag 18
|||||||||||||||
Db 72 TCAGGTGCAGGTCAG 58

RESULT 29
BG100182/c
LOCUS BG100182 306 bp mRNA EST 29-JAN-2001
DEFINITION ux94g04.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA clone IMAGE:3656262 5', mRNA sequence.
ACCESSION BG100182
VERSION BG100182.1 GI:12595499
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 306)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1418566
Seq primer: Primer name ambiguous.
FEATURES Location/Qualifiers
source 1. .306

```
                     /organism="Mus musculus"
                     /strain="CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3656262"
                     /clone_lib="McCarrey Eddy type B spermatogonia"
                     /sex="male"
                     /tissue_type="type B spermatogonia, pooled from multiple
                     mice"
                     /dev_stage="8 day"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
                     ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
                     [5'-(GA)10-ACTAGTCTCGAGTTTTTTTTTTTT-3'] and directionally
                     cloned  using 5' linkers 5'-AATTCGGCACGAG-3' and
                     5'-CTCGTGCCG-3'. Size selection of >400bp material gives
                     average insert size ranging from 1-2 kb. Library was mass
                     excised (from  lambda-UniZAP-XR) and resulting
                     single-stranded phagemids were prepped and tranformed
                     into DH10B. Library contains 96% recombinants.
                     References: J. Androl. 20:635-639 and Gene 25:263-269.
                     Library constructed and donated by J. McCarrey, Ph.D.
                     (Southwest Foundation for Biomedical Research, Dept. of
                     Genetics); excision done by E.M. Eddy, Ph.D. (National
                     Institutes of  Health, National Institute of Environmental
                     Health Sciences).  Original lambda-based library is
                     available through ATCC, catalog #63417."
BASE COUNT       62 a     82 c     93 g     69 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 173;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           ||||||||||||||||
Db      95 GTCAGGTGCAGGTCA 81


RESULT  30
AI138808
LOCUS       AI138808      382 bp    mRNA            EST       28-OCT-1998
DEFINITION  qd98g02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1737554
            3', mRNA sequence.
ACCESSION   AI138808
VERSION     AI138808.1  GI:3644780
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 593   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 362.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1737554"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      103 a     84 c     75 g    120 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 16;  Length 382;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagca 20
           ||||||||||||||||
Db     311 AGGTGCAGGTCAGCA 325


RESULT  31
AW522699/c
LOCUS       AW522699      385 bp    mRNA            EST       06-MAR-2000
DEFINITION  UI-R-BO0-ahk-c-01-0-UI.s1 UI-R-BO0 Rattus norvegicus cDNA clone
            UI-R-BO0-ahk-c-01-0-UI 3', mRNA sequence.
ACCESSION   AW522699
VERSION     AW522699.1  GI:7165124
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized medulla library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO0-ahk-c-01-0-UI"
                     /clone_lib="UI-R-BO0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
```

(UI-R-BO0) is a subtracted library derived from a mixture of the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BO0
TAG_TISSUE=medulla
TAG_SEQ=GAACCG"
BASE COUNT 87 a 104 c 98 g 96 t
ORIGIN

Query Match 60.0%; Score 15; DB 117; Length 385;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gtgcaggtcagcacg 22
||||||||||||||
Db 336 GTGCAGGTCAGCACG 322

RESULT 32
AW486040/c
LOCUS AW486040 396 bp mRNA EST 10-JUL-2000
DEFINITION 69428 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW486040
VERSION AW486040.1 GI:7056146
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 396)
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid ,W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 45 row: I column: 7
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .396
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 20 and day 40 embryos."
BASE COUNT 70 a 134 c 127 g 65 t
ORIGIN

Query Match 60.0%; Score 15; DB 116; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgcaggtcagc 19
|||||||||||||||
Db 221 CAGGTGCAGGTCAGC 207

RESULT 33
AA062258/c
LOCUS AA062258 407 bp mRNA EST 03-FEB-1997
DEFINITION ml55g08.r1 Stratagene mouse testis (#937308) Mus musculus cDNA clone IMAGE:515966 5', mRNA sequence.
ACCESSION AA062258
VERSION AA062258.1 GI:1556057
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 407)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:309814
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 387.
FEATURES Location/Qualifiers
source 1. .407
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:515966"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 88 a 117 c 123 g 79 t
ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gtcaggtgcaggtca 17
|||||||||||||||
Db 15 GTCAGGTGCAGGTCA 1

RESULT 34
BF599674/c
LOCUS BF599674 407 bp mRNA EST 13-DEC-2000
DEFINITION 263546 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

```
ACCESSION   BF599674
VERSION     BF599674.1  GI:11696393
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 35  row: E  column: 3
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       69 a    139 c    135 g     64 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 151;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db      93 CAGGTGCAGGTCAGC 79


RESULT  35
AQ085001/c
LOCUS       AQ085001      408 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2269_A2_B10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2269 Col=20 Row=C, DNA sequence.
ACCESSION   AQ085001
VERSION     AQ085001.1  GI:3454218
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2269  row: C  column: 20
            Class: BAC ends
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2269 Col=20 Row=C"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      104 a    111 c     78 g    115 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 224;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     396 CAGGTGCAGGTCAGC 382


RESULT  36
BF653430/c
LOCUS       BF653430      409 bp    mRNA            EST       20-DEC-2000
DEFINITION  277257 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF653430
VERSION     BF653430.1  GI:11918562
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 69  row: P  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
```

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendonosus muscle, and fetal longissimus muscle."
BASE COUNT 80 a 138 c 129 g 62 t
ORIGIN

Query Match 60.0%; Score 15; DB 151; Length 409;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 caggtgcaggtcagc 19
||||||||||||||||
Db 255 CAGGTGCAGGTCAGC 241

RESULT 37
AA146146/c
LOCUS AA146146 416 bp mRNA EST 18-FEB-1997
DEFINITION mq65d12.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:583607 5', mRNA sequence.
ACCESSION AA146146
VERSION AA146146.1 GI:1715582
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 416)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:358255
Seq primer: -28M13 rev2 from Amersham.
FEATURES Location/Qualifiers
source 1. .416
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:583607"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 129 a 83 c 125 g 79 t
ORIGIN

Query Match 60.0%; Score 15; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gtgcaggtcagcacg 22
|||||||||||||||
Db 173 GTGCAGGTCAGCACG 159

RESULT 38
AA881231/c
LOCUS AA881231 446 bp mRNA EST 26-MAR-1998
DEFINITION vx10d04.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:1264039 5', mRNA sequence.
ACCESSION AA881231
VERSION AA881231.1 GI:2990541
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 446)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:666591
Seq primer: -28m13 rev2 ET from Amersham.
FEATURES Location/Qualifiers
source 1. .446
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1264039"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 141 a 87 c 131 g 87 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 446;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gtgcaggtcagcacg 22
|||||||||||||||
Db 158 GTGCAGGTCAGCACG 144

```
RESULT  39
BF604067/c
LOCUS       BF604067      457 bp    mRNA            EST       13-DEC-2000
DEFINITION  269755 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF604067
VERSION     BF604067.1  GI:11702307
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 50  row: I  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       82 a    152 c    142 g     80 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 151;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     215 CAGGTGCAGGTCAGC 201


RESULT  40
AQ921326/c
LOCUS       AQ921326      477 bp    DNA             GSS       21-DEC-1999
DEFINITION  RPCI-23-273P15.TV RPCI-23 Mus musculus genomic clone RPCI-23-273P15
            , DNA sequence.
ACCESSION   AQ921326
VERSION     AQ921326.1  GI:6610329
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 273  row: P  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-273P15"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      101 a    107 c    125 g    144 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 235;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      68 GTCAGGTGCAGGTCA 54


RESULT  41
BE808281/c
LOCUS       BE808281      492 bp    mRNA            EST       20-SEP-2000
DEFINITION  213447 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE808281
VERSION     BE808281.1  GI:10239393
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
```

```
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 68  row: P  column: 2
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .492
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 2BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from testis, thymus,
                      semitendonosus muscle, longissimus muscle, pancreas,
                      adrenal, and endometrium."
BASE COUNT       96 a     161 c     149 g      85 t       1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 140;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     263 CAGGTGCAGGTCAGC 249


RESULT  42
AZ261270/c
LOCUS       AZ261270      495 bp    DNA             GSS       26-JUL-2000
DEFINITION  RPCI-23-440I8.TV RPCI-23 Mus musculus genomic clone RPCI-23-440I8,
            DNA sequence.
ACCESSION   AZ261270
VERSION     AZ261270.1  GI:9469499
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-440I8.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 440  row: I  column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .495
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-23-440I8"
                      /clone_lib="RPCI-23"
                      /sex="Female"
                      /lab_host="DH10B"
                      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                      EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                      brain genomic DNA was isolated and  partially digested
                      with a combination of EcoRI and EcoRI Methylase.   Size
                      selected DNA was cloned into the pBACe3.6 vector at the
                      EcoRI  sites.  The ligation products were transformed into
                      DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      106 a     110 c     132 g     147 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 240;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      66 GTCAGGTGCAGGTCA 52


RESULT  43
BF652819/c
LOCUS       BF652819      496 bp    mRNA            EST       20-DEC-2000
DEFINITION  276504 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF652819
VERSION     BF652819.1  GI:11917951
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 67  row: C  column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .496
                      /organism="Bos taurus"
                      /db_xref="taxon:9913"
                      /clone_lib="MARC 3BOV"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from marrow, alveolar
                      macrophage, ovary, fetal semitendonosus muscle, and fetal
                      longissimus muscle."
BASE COUNT      100 a     164 c     148 g      84 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 151;  Length 496;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
```

```
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     290 CAGGTGCAGGTCAGC 276


RESULT  44
BF567735
LOCUS       BF567735      525 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-BO0-ahk-c-01-0-UI.r1 UI-R-BO0 Rattus norvegicus cDNA clone
            UI-R-BO0-ahk-c-01-0-UI 5', mRNA sequence.
ACCESSION   BF567735
VERSION     BF567735.1  GI:11677465
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1796684
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO0-ahk-c-01-0-UI"
                     /clone_lib="UI-R-BO0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     (UI-R-BO0) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)"
BASE COUNT      131 a    143 c    138 g    112 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 150;  Length 525;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gtgcaggtcagcacg 22
           |||||||||||||||
Db     491 GTGCAGGTCAGCACG 505


RESULT  45
AW290780/c
LOCUS       AW290780      529 bp    mRNA            EST       16-JAN-2000
DEFINITION  NXNV046H01F Nsf Xylem Normal wood Vertical Pinus taeda cDNA clone
            NXNV046H01 5', mRNA sequence.
ACCESSION   AW290780
VERSION     AW290780.1  GI:6697416
KEYWORDS    EST.
SOURCE      loblolly pine.
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Sederoff,R.
  TITLE     Molecular Basis of Wood Formation in the Pine Megagenome
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Johnson, Arthur
            North Carolina State University
            Tel: 919 515 7800
            Fax: 919 515 7801
            Email: ajohnson@unity.ncsu.edu
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Pinus taeda"
                     /db_xref="taxon:3352"
                     /clone="NXNV046H01"
                     /clone_lib="Nsf Xylem Normal wood Vertical"
                     /note="Vector: BlueScript SK; Site_1: Eco RI; The
                     sequences contain a 'cDNA adapter' between the EcoRI site
                     and the start of the EST. The adapter sequence is
                     'AATTCGGCACGAG'."
BASE COUNT      157 a     89 c    129 g    133 t     21 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 114;  Length 529;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcacgt 23
           |||||||||||||||
Db      65 TGCAGGTCAGCACGT 51


Search completed: October  9, 2001, 15:15:40
Job time: 13662 sec
```

09/396,196

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:43:40 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 208678

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 BioB gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

	9	25	100.0	5872	9	A93679 A93679 Sequence 6
	10	25	100.0	5872	9	AR101809 AR101809 Sequence
	11	25	100.0	5872	9	AR101810 AR101810 Sequence
	12	25	100.0	11022	1	AE000180 AE000180 Escherich
	13	25	100.0	13501	1	AE005258 AE005258 Escherich
	14	25	100.0	297816	2	AP002553 AP002553 Escherich
	15	24	96.0	5526	2	AF250776 AF250776 Unculture
c	16	17	68.0	25256	91	AP000954 AP000954 Homo sapi
c	17	17	68.0	32767	77	AC087305 AC087305 Homo sapi
	18	17	68.0	60072	62	AC011937 AC011937 Homo sapi
c	19	17	68.0	86757	84	HS54J18 AL133366 Homo sapi
c	20	17	68.0	115551	90	AL353582 AL353582 Human DNA
c	21	17	68.0	147706	12	AC083945 AC083945 Genomic S
	22	17	68.0	155172	66	AC021110 AC021110 Homo sapi
c	23	17	68.0	174815	82	AP001390 AP001390 Homo sapi
c	24	17	68.0	228516	85	AC002523 AC002523 Homo sapi
c	25	17	68.0	340000	91	AP001690 AP001690 Homo sapi
c	26	16	64.0	605	93	HSAPAF20 AF098889 Homo sapi
c	27	16	64.0	1829	54	G07181 G07181 human STS W
c	28	16	64.0	4602	91	D13644 D13644 Human mRNA
	29	16	64.0	5469	3	LLA012388 AJ012388 Lactococc
c	30	16	64.0	43741	90	AL161800 AL161800 Human DNA
c	31	16	64.0	68894	75	AC074073 AC074073 Homo sapi
c	32	16	64.0	105937	12	AC012679 AC012679 Arabidops
	33	16	64.0	110525	85	AB020878 AB020878 Homo sapi
	34	16	64.0	111312	87	AC011248 AC011248 Homo sapi
	35	16	64.0	118636	82	AL590387 AL590387 Homo sapi
	36	16	64.0	129010	90	AL159168 AL159168 Human DNA
c	37	16	64.0	139218	65	AC019289 AC019289 Homo sapi
c	38	16	64.0	152452	77	AC090105 AC090105 Homo sapi
c	39	16	64.0	153221	73	AC067998 AC067998 Homo sapi
c	40	16	64.0	155211	65	AC019258 AC019258 Homo sapi
c	41	16	64.0	156184	71	AC041018 AC041018 Homo sapi
	42	16	64.0	158408	90	AL391821 AL391821 Human DNA
	43	16	64.0	167110	67	AC022621 AC022621 Homo sapi
	44	16	64.0	167176	68	AC024560 AC024560 Homo sapi
	45	16	64.0	167803	61	AC010816 AC010816 Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   3
A11530
LOCUS       A11530       1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
```

```
Db      47 ATTGTCGCAAGTCACAGAATTATTT 71


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 10;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      65 ATTGTCGCAAGTCACAGAATTATTT 89


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 2;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db   2035 ATTGTCGCAAGTCACAGAATTATTT 2059


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
```

```
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t

ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
```

```
                    YINLAFVDASTLGFTSRL"
BASE COUNT      1318 a    1552 c    1695 g    1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  10
AR101809
LOCUS       AR101809      5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT      1318 a    1552 c    1695 g    1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  11
AR101810
LOCUS       AR101810      5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT      1318 a    1552 c    1695 g    1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
```

```
                     /db_xref="taxon:83333"
repeat_region        <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
gene                 complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
CDS                  complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
                     /protein_id="AAC73859.1"
                     /db_xref="GI:1786989"
                     /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                     LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                     AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                     DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                     LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                     NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                     AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                     DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter             complement(1507. .1535)
                     /note="factor Sigma70; predicted +1 start at 806574"
gene                 complement(1582. .2058)
                     /gene="ybhB"
                     /note="b0773"
CDS                  complement(1582. .2058)
                     /gene="ybhB"
                     /function="orf; Unknown"
                     /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC73860.1"
                     /db_xref="GI:1786990"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind         complement(2108. .2124)
                     /note="central position to predicted promoter:85.5"
                     /bound_moiety="RhaS predicted site"
gene                 complement(2117. .3406)
                     /gene="bioA"
                     /note="b0774"
CDS                  complement(2117. .3406)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diaminopelargonic acid synthetase"
                     /protein_id="AAC73861.1"
                     /db_xref="GI:1786991"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter             complement(2193. .2221)
                     /gene="bioA"
                     /note="factor Sigma70; predicted +1 start at 807260"
protein_bind         3411. .3450
                     /note="central position to bioB promoter: -20"
                     /bound_moiety="BioB documented site"
protein_bind         3411. .3450
                     /note="central position to predicted promoter: -20;genetic
                     evidence for the site"
                     /bound_moiety="BioB predicted site"
protein_bind         complement(3411. .3450)
                     /note="central position to predicted promoter:50"
                     /bound_moiety="BioB predicted site"
protein_bind         complement(3411. .3450)
                     /note="central position to bioA promoter:50"
                     /bound_moiety="BioB documented site"
promoter             3413. .3441
                     /note="factor Sigma70; promoter bioB; documented +1 at
                     808525"
promoter             complement(3447. .3473)
                     /note="factor Sigma70; promoter bioA; documented +1 at
                     808515"
gene                 3493. .4533
                     /gene="bioB"
                     /note="b0775"
CDS                  3493. .4533
                     /gene="bioB"
                     /EC_number="2.8.1.-"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis, sulfur insertion?"
                     /protein_id="AAC73862.1"
                     /db_xref="GI:1786992"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
gene                 4530. .5684
                     /gene="bioF"
                     /note="b0776"
CDS                  4530. .5684
                     /gene="bioF"
                     /EC_number="2.3.1.47"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAC73863.1"
                     /db_xref="GI:1786993"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL

  Query Match          100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    3516 ATTGTCGCAAGTCACAGAATTATTT 3540


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
```

```
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
```

```
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
          gene       6462. .7160
                     /gene="Z0990"
          CDS        6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
          gene       complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
          CDS        complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
          gene       complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
          CDS        complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    9602 ATTGTCGCAAGTCACAGAATTATTT 9626


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
          gene       79. .1245
                     /gene="ECs0753"
          CDS        79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
```

```
gene            1245. .2114
                /gene="ECs0754"
CDS             1245. .2114
                /gene="ECs0754"
                /note="similar to SUCD_ECOLI gi|1786949 percent identity
                100 in 289 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="succinyl-CoA synthetase alpha subunit"
                /protein_id="BAB34177.1"
                /db_xref="GI:13360213"
                /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
gene            2781. .3686
                /gene="ECs0755"
CDS             2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

```
 Query Match             100.0%;  Score 25;  DB 2;  Length 297816;
 Best Local Similarity   100.0%;  Pred. No. 3.9e-05;
 Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db   92334 ATTGTCGCAAGTCACAGAATTATTT 92358
```

```
RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
```

```
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             96.0%;  Score 24;  DB 2;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.00021;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    1987 TTGTCGCAAGTCACAGAATTATTT 2010


RESULT  16
AP000954/c
LOCUS       AP000954    25256 bp    DNA             PRI       27-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B705D10,
            LL56-APP region, complete sequence.
ACCESSION   AP000954
VERSION     AP000954.2  GI:6778729
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:B705D10.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 25256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 25,256 genomic DNA of 21q21.1-q21.2
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 25256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Jan 27, 2000 this sequence version replaced gi:6630654.
```

```
           Sequence updated (26-Jan-2000).
FEATURES             Location/Qualifiers
     source          1. .25256
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="B705D10"
                     /map="21q21.1-q21.2"
BASE COUNT     8452 a   4090 c   4292 g   8422 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 91;  Length 25256;
  Best Local Similarity   100.0%;  Pred. No. 3.5;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db    1714 AAGTCACAGAATTATTT 1698


RESULT  17
AC087305/c
LOCUS       AC087305    32767 bp    DNA             HTG       23-DEC-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC087305
VERSION     AC087305.1  GI:11990696
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-238P13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12071
                Center clone name: 238_P_13
            --------------.
            * NOTE: This record contains 40 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      707: contig of 707 bp in length
            *      708 807: gap of      100 bp
            *      808     1527: contig of 720 bp in length
            *     1528 1627: gap of      100 bp
            *     1628     2348: contig of 721 bp in length
            *     2349 2448: gap of      100 bp
            *     2449     3157: contig of 709 bp in length
            *     3158 3257: gap of      100 bp
            *     3258     3973: contig of 716 bp in length
            *     3974 4073: gap of      100 bp
            *     4074     4799: contig of 726 bp in length
            *     4800 4899: gap of      100 bp
            *     4900     5617: contig of 718 bp in length
            *     5618 5717: gap of      100 bp
            *     5718     6450: contig of 733 bp in length
            *     6451 6550: gap of      100 bp
            *     6551     7259: contig of 709 bp in length
            *     7260 7359: gap of      100 bp
            *     7360     8066: contig of 707 bp in length
            *     8067 8166: gap of      100 bp
            *     8167     8907: contig of 741 bp in length
            *     8908 9007: gap of      100 bp
            *     9008     9716: contig of 709 bp in length
            *     9717 9816: gap of      100 bp
            *     9817    10532: contig of 716 bp in length
            *    10533 10632: gap of      100 bp
            *    10633    11354: contig of 722 bp in length
            *    11355 11454: gap of      100 bp
            *    11455    12166: contig of 712 bp in length
            *    12167 12266: gap of      100 bp
            *    12267    12976: contig of 710 bp in length
            *    12977 13076: gap of      100 bp
            *    13077    13789: contig of 713 bp in length
            *    13790 13889: gap of      100 bp
            *    13890    14617: contig of 728 bp in length
            *    14618 14717: gap of      100 bp
            *    14718    15452: contig of 735 bp in length
            *    15453 15552: gap of      100 bp
            *    15553    16276: contig of 724 bp in length
            *    16277 16376: gap of      100 bp
            *    16377    17111: contig of 735 bp in length
            *    17112 17211: gap of      100 bp
            *    17212    17925: contig of 714 bp in length
            *    17926 18025: gap of      100 bp
            *    18026    18746: contig of 721 bp in length
            *    18747 18846: gap of      100 bp
            *    18847    19569: contig of 723 bp in length
            *    19570 19669: gap of      100 bp
            *    19670    20384: contig of 715 bp in length
            *    20385 20484: gap of      100 bp
            *    20485    21190: contig of 706 bp in length
            *    21191 21290: gap of      100 bp
            *    21291    22028: contig of 738 bp in length
            *    22029 22128: gap of      100 bp
            *    22129    22841: contig of 713 bp in length
            *    22842 22941: gap of      100 bp
            *    22942    23653: contig of 712 bp in length
            *    23654 23753: gap of      100 bp
            *    23754    24478: contig of 725 bp in length
            *    24479 24578: gap of      100 bp
            *    24579    25305: contig of 727 bp in length
```

```
            *    25306 25405: gap of      100 bp
            *    25406    26143: contig of 738 bp in length
            *    26144 26243: gap of      100 bp
            *    26244    26968: contig of 725 bp in length
            *    26969 27068: gap of      100 bp
            *    27069    27785: contig of 717 bp in length
            *    27786 27885: gap of      100 bp
            *    27886    28584: contig of 699 bp in length
            *    28585 28684: gap of      100 bp
            *    28685    29467: contig of 783 bp in length
            *    29468 29567: gap of      100 bp
            *    29568    30292: contig of 725 bp in length
            *    30293 30392: gap of      100 bp
            *    30393    31117: contig of 725 bp in length
            *    31118 31217: gap of      100 bp
            *    31218    31944: contig of 727 bp in length
            *    31945 32044: gap of      100 bp
            *    32045    32767: contig of 723 bp in length.
FEATURES             Location/Qualifiers
     source          1. .32767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-238P13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     8045 a   5670 c   5396 g   9330 t   4326 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 77;  Length 32767;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db   30730 GCAAGTCACAGAATTAT 30714


RESULT  18
AC011937
LOCUS       AC011937    60072 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-16E13, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC011937
VERSION     AC011937.2  GI:7144927
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60072)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60072)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6056246.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L3522
    Center clone name: 16_E_13
--------------.
* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*         1      763: contig of 763 bp in length
*       764 863: gap of      100 bp
*       864     1659: contig of 796 bp in length
*      1660 1759: gap of      100 bp
*      1760     2606: contig of 847 bp in length
*      2607 2706: gap of      100 bp
*      2707     3494: contig of 788 bp in length
*      3495 3594: gap of      100 bp
*      3595     4391: contig of 797 bp in length
*      4392 4491: gap of      100 bp
*      4492     5168: contig of 677 bp in length
*      5169 5268: gap of      100 bp
*      5269     6058: contig of 790 bp in length
*      6059 6158: gap of      100 bp
*      6159     6968: contig of 810 bp in length
*      6969 7068: gap of      100 bp
*      7069     7865: contig of 797 bp in length
*      7866 7965: gap of      100 bp
*      7966     8753: contig of 788 bp in length
*      8754 8853: gap of      100 bp
*      8854     9663: contig of 810 bp in length
*      9664 9763: gap of      100 bp
*      9764    10563: contig of 800 bp in length
*     10564 10663: gap of      100 bp
*     10664    11465: contig of 802 bp in length
*     11466 11565: gap of      100 bp
*     11566    12365: contig of 800 bp in length
*     12366 12465: gap of      100 bp
*     12466    13274: contig of 809 bp in length
*     13275 13374: gap of      100 bp
*     13375    14163: contig of 789 bp in length
*     14164 14263: gap of      100 bp
*     14264    15064: contig of 801 bp in length
*     15065 15164: gap of      100 bp
*     15165    15965: contig of 801 bp in length
*     15966 16065: gap of      100 bp
*     16066    16852: contig of 787 bp in length
*     16853 16952: gap of      100 bp
*     16953    17775: contig of 823 bp in length
*     17776 17875: gap of      100 bp
*     17876    18674: contig of 799 bp in length
*     18675 18774: gap of      100 bp
*     18775    19477: contig of 703 bp in length
*     19478 19577: gap of      100 bp
*     19578    20375: contig of 798 bp in length
*     20376 20475: gap of      100 bp
*     20476    21252: contig of 777 bp in length
*     21253 21352: gap of      100 bp
*     21353    22153: contig of 801 bp in length
*     22154 22253: gap of      100 bp
```

```
*    22254    23063: contig of 810 bp in length
*    23064 23163: gap of      100 bp
*    23164    23970: contig of 807 bp in length
*    23971 24070: gap of      100 bp
*    24071    24697: contig of 627 bp in length
*    24698 24797: gap of      100 bp
*    24798    25395: contig of 598 bp in length
*    25396 25495: gap of      100 bp
*    25496    26303: contig of 808 bp in length
*    26304 26403: gap of      100 bp
*    26404    27193: contig of 790 bp in length
*    27194 27293: gap of      100 bp
*    27294    27964: contig of 671 bp in length
*    27965 28064: gap of      100 bp
*    28065    28724: contig of 660 bp in length
*    28725 28824: gap of      100 bp
*    28825    29615: contig of 791 bp in length
*    29616 29715: gap of      100 bp
*    29716    30499: contig of 784 bp in length
*    30500 30599: gap of      100 bp
*    30600    31391: contig of 792 bp in length
*    31392 31491: gap of      100 bp
*    31492    32268: contig of 777 bp in length
*    32269 32368: gap of      100 bp
*    32369    33177: contig of 809 bp in length
*    33178 33277: gap of      100 bp
*    33278    34070: contig of 793 bp in length
*    34071 34170: gap of      100 bp
*    34171    34978: contig of 808 bp in length
*    34979 35078: gap of      100 bp
*    35079    35870: contig of 792 bp in length
*    35871 35970: gap of      100 bp
*    35971    36751: contig of 781 bp in length
*    36752 36851: gap of      100 bp
*    36852    37644: contig of 793 bp in length
*    37645 37744: gap of      100 bp
*    37745    38573: contig of 829 bp in length
*    38574 38673: gap of      100 bp
*    38674    39462: contig of 789 bp in length
*    39463 39562: gap of      100 bp
*    39563    40367: contig of 805 bp in length
*    40368 40467: gap of      100 bp
*    40468    41259: contig of 792 bp in length
*    41260 41359: gap of      100 bp
*    41360    42171: contig of 812 bp in length
*    42172 42271: gap of      100 bp
*    42272    43084: contig of 813 bp in length
*    43085 43184: gap of      100 bp
*    43185    43981: contig of 797 bp in length
*    43982 44081: gap of      100 bp
*    44082    44891: contig of 810 bp in length
*    44892 44991: gap of      100 bp
*    44992    45788: contig of 797 bp in length
*    45789 45888: gap of      100 bp
*    45889    46671: contig of 783 bp in length
*    46672 46771: gap of      100 bp
*    46772    47570: contig of 799 bp in length
*    47571 47670: gap of      100 bp
*    47671    48460: contig of 790 bp in length
*    48461 48560: gap of      100 bp
*    48561    49352: contig of 792 bp in length
*    49353 49452: gap of      100 bp
*    49453    50256: contig of 804 bp in length
*    50257 50356: gap of      100 bp
*    50357    51147: contig of 791 bp in length
*    51148 51247: gap of      100 bp
*    51248    52076: contig of 829 bp in length
*    52077 52176: gap of      100 bp
*    52177    52976: contig of 800 bp in length
*    52977 53076: gap of      100 bp
*    53077    53891: contig of 815 bp in length
*    53892 53991: gap of      100 bp
*    53992    54778: contig of 787 bp in length
*    54779 54878: gap of      100 bp
*    54879    55683: contig of 805 bp in length
*    55684 55783: gap of      100 bp
*    55784    56484: contig of 701 bp in length
*    56485 56584: gap of      100 bp
*    56585    57369: contig of 785 bp in length
*    57370 57469: gap of      100 bp
*    57470    58276: contig of 807 bp in length
*    58277 58376: gap of      100 bp
*    58377    59169: contig of 793 bp in length
*    59170 59269: gap of      100 bp
*    59270    60072: contig of 803 bp in length.
FEATURES             Location/Qualifiers
     source          1. .60072
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-16E13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    15923 a   9785 c  10180 g  17174 t   7010 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 62;  Length 60072;
  Best Local Similarity   100.0%;  Pred. No. 3.3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db    9213 AAGTCACAGAATTATTT 9229


RESULT  19
HS54J18/c
LOCUS       HS54J18     86757 bp    DNA             HTG       28-SEP-2000
DEFINITION  Homo sapiens chromosome 21 clone RPCIP704J1854 map 21q21, ***
            SEQUENCING IN PROGRESS ***, 18 unordered pieces.
ACCESSION   AL133366
VERSION     AL133366.2  GI:10432568
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86757)
  AUTHORS   Nordsiek,G., Conrad,A., Dose,S., Grimm,M., Hildmann,T.,
            Hornischer,K., Loehnert,T.H., Scharfe,M., Schoen,O., Yaspo,M.L. and
            Bloecker,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) GBF, Dept. of Genome Analysis, Mascheroder
            Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
            nordsiek@gbf.de, bloecker@gbf.de
COMMENT     On Sep 29, 2000 this sequence version replaced gi:6981908.
            Collaborators:
            GBF, Dept. of Genome Analysis
            Mascheroder Weg 1, D-38124 Braunschweig, Germany
            and
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany
            All annotations in this database entry are developed by
            computational tools. It is therefore not explicitly noted in the
            feature lines that evidence is not experimental.
            PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            +   Analysis and annotation were performed with the automatic    +
            +          'first-pass' annotation and submission tool           +
            +             'AnnoMitter' (Hornischer & Bloecker).              +
            +                 Programs used by 'AnnoMitter':                 +
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            > 'Tandem Repeats': GDE 2.2 option 'tandem'
            .   Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
              Treat N's as mismatches? YES; Allow uniform consensi? NO >
            'Inverted Repeats': GDE 2.2 option 'inverted'
            > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
```

```
            Islands': GDE 2.2 option 'cpg'
            .   CpG island region size 100 bp;
            .   Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
           e-PCR (Schuler)
            .   Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
           STS database: 'dbSTS markers'.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13423: contig of 13423 bp in length
            *    13424 13523: gap of      100 bp
            *    13524    15096: contig of 1573 bp in length
            *    15097 15196: gap of      100 bp
            *    15197    16448: contig of 1252 bp in length
            *    16449 16548: gap of      100 bp
            *    16549    19018: contig of 2470 bp in length
            *    19019 19118: gap of      100 bp
            *    19119    24811: contig of 5693 bp in length
            *    24812 24911: gap of      100 bp
            *    24912    34477: contig of 9566 bp in length
            *    34478 34577: gap of      100 bp
            *    34578    37261: contig of 2684 bp in length
            *    37262 37361: gap of      100 bp
            *    37362    41884: contig of 4523 bp in length
            *    41885 41984: gap of      100 bp
            *    41985    45057: contig of 3073 bp in length
            *    45058 45157: gap of      100 bp
            *    45158    46158: contig of 1001 bp in length
            *    46159 46258: gap of      100 bp
            *    46259    50846: contig of 4588 bp in length
            *    50847 50946: gap of      100 bp
            *    50947    53362: contig of 2416 bp in length
            *    53363 53462: gap of      100 bp
            *    53463    55903: contig of 2441 bp in length
            *    55904 56003: gap of      100 bp
            *    56004    64319: contig of 8316 bp in length
            *    64320 64419: gap of      100 bp
            *    64420    67502: contig of 3083 bp in length
            *    67503 67602: gap of      100 bp
            *    67603    71620: contig of 4018 bp in length
            *    71621 71720: gap of      100 bp
            *    71721    74980: contig of 3260 bp in length
            *    74981 75080: gap of      100 bp
            *    75081    86757: contig of 11677 bp in length.
FEATURES             Location/Qualifiers
     source          1. .86757
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21"
                     /clone="RPCIP704J1854"
     repeat_region   4088. .4150
                     /note="homology = 73.0%, counts = 7"
                     /rpt_family="tttttatta repeat"
                     /rpt_type=TANDEM
     repeat_region   6744. .6807
                     /note="homology = 78.1%, counts = 8"
                     /rpt_family="tttttatt repeat"
                     /rpt_type=TANDEM
     satellite       6766. .6792
                     /note="TTTA repeat"
     satellite       6793. .6804
                     /note="TAT repeat"
     satellite       7177. .7192
                     /note="AACAA repeat"
     satellite       7426. .7440
                     /note="AATTA repeat"
     repeat_region   8984. .9033
                     /note="IR1, 92% complementary to IR1' (10757. .10806)"
                     /rpt_type=INVERTED
     satellite       9075. .9094
                     /note="AAAC repeat"
     repeat_region   10757. .10806
                     /note="IR1', 92% complementary to IR1 (8984. .9033)"
                     /rpt_type=INVERTED
     satellite       12276. .12290
                     /note="TTG repeat"
     satellite       13116. .13129
                     /note="ATAA repeat"
     repeat_region   15197. .15238
                     /note="IR2, 97% complementary to IR2' (15480. .15521)"
                     /rpt_type=INVERTED
     repeat_region   15480. .15521
                     /note="IR2', 97% complementary to IR2 (15197. .15238)"
                     /rpt_type=INVERTED
     satellite       22885. .22929
                     /note="TG repeat"
     repeat_region   22885. .22928
                     /note="homology = 100.0%, counts = 22"
                     /rpt_family="tg repeat"
                     /rpt_type=TANDEM
     repeat_region   25803. .25870
                     /note="homology = 88.2%, counts = 34"
                     /rpt_family="at repeat"
                     /rpt_type=TANDEM
     repeat_region   25819. .25843
                     /note="IR3, 92% complementary to IR3' (25846. .25870)"
                     /rpt_type=INVERTED
     satellite       25828. .25870
                     /note="TA repeat"
     repeat_region   25846. .25870
                     /note="IR3', 92% complementary to IR3 (25819. .25843)"
                     /rpt_type=INVERTED
     satellite       26732. .26745
                     /note="CAA repeat"
     satellite       35152. .35163
                     /note="TC repeat"
     satellite       38666. .38678
                     /note="TG repeat"
     satellite       40287. .40299
                     /note="ATTT repeat"
     repeat_region   42411. .42436
                     /note="IR4, 84% complementary to IR4' (44122. .44147)"
                     /rpt_type=INVERTED
     repeat_region   44116. .44245
                     /note="homology = 63.1%, counts = 10"
                     /rpt_family="aaaaaaaaaaaat repeat"
                     /rpt_type=TANDEM
     repeat_region   44122. .44147
                     /note="IR4', 84% complementary to IR4 (42411. .42436)"
                     /rpt_type=INVERTED
     satellite       44370. .44383
                     /note="AAAAT repeat"
     repeat_region   46995. .47082
                     /note="homology = 68.2%, counts = 8"
                     /rpt_family="gaaaaaaaaga repeat"
                     /rpt_type=TANDEM
     repeat_region   47093. .47127
                     /note="IR5, 82% complementary to IR5' (47138. .47173)"
                     /rpt_type=INVERTED
     repeat_region   47114. .47175
                     /note="homology = 85.5%, counts = 31"
                     /rpt_family="ta repeat"
                     /rpt_type=TANDEM
     repeat_region   47138. .47173
                     /note="IR5', 82% complementary to IR5 (47093. .47127)"
                     /rpt_type=INVERTED
     satellite       47159. .47175
                     /note="AT repeat"
     repeat_region   47864. .47953
                     /note="homology = 66.7%, counts = 10"
```

```
                     /rpt_family="agaaaaaaa repeat"
                     /rpt_type=TANDEM
     repeat_region   48093. .48128
                     /note="homology = 83.3%, counts = 18"
                     /rpt_family="tc repeat"
                     /rpt_type=TANDEM
     repeat_region   48328. .48366
                     /note="IR6, 82% complementary to IR6' (49401. .49440)"
                     /rpt_type=INVERTED
     repeat_region   49049. .49084
                     /note="homology = 88.9%, counts = 9"
                     /rpt_family="aaat repeat"
                     /rpt_type=TANDEM
     satellite       49049. .49069

  Query Match             68.0%;  Score 17;  DB 84;  Length 86757;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db   65420 AAGTCACAGAATTATTT 65404


RESULT  20
AL353582/c
LOCUS       AL353582   115551 bp    DNA             PRI       22-MAR-2001
DEFINITION  Human DNA sequence from clone CTD-2190D5 on chromosome 20 Contains
            ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL353582
VERSION     AL353582.16  GI:10443412
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115551)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:9988285.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            CTD-2190D5 is from the CalTech genomic sperm BAC library D. VECTOR:
            pBeloBAC11
            IMPORTANT: This sequence is not the entire insert of clone
            CTD-2190D5 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-568C11 is at 115452 in this
            sequence. The true right end of clone RP5-966J20 is at 100 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .115551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="CTD-2190D5"
                     /clone_lib="CIT-HSP-D1"
     repeat_region   1. .260
                     /note="10 copies 26 mer 88% conserved"
     repeat_region   1400. .1691
                     /note="AluSg repeat: matches 1. .292 of consensus"
     repeat_region   1692. .1818
                     /note="L1PA16 repeat: matches 6020. .6157 of consensus"
     repeat_region   1878. .2076
                     /note="L1MC4 repeat: matches 7699. .7977 of consensus"
     repeat_region   2719. .3328
                     /note="305 copies 2 mer cc 52% conserved"
     repeat_region   2720. .2777
                     /note="2 copies 29 mer 94% conserved"
     repeat_region   2905. .3078
                     /note="2 copies 87 mer 92% conserved"
     repeat_region   2955. .3494
                     /note="6 copies 90 mer 85% conserved"
     repeat_region   2985. .3484
                     /note="50 copies 10 mer ccccctcctg 60% conserved"
     repeat_region   3001. .3510
                     /note="17 copies 30 mer 89% conserved"
     repeat_region   3496. .3625
                     /note="13 copies 10 mer cccctttcctg 63% conserved"
     repeat_region   3512. .3631
                     /note="4 copies 30 mer 87% conserved"
     repeat_region   4283. .4377
                     /note="MER5A repeat: matches 63. .167 of consensus"
     repeat_region   4459. .4588
                     /note="65 copies 2 mer tg 77% conserved"
     repeat_region   4460. .4579
                     /note="4 copies 30 mer 80% conserved"
     repeat_region   4591. .4800
                     /note="7 copies 30 mer 77% conserved"
     repeat_region   4592. .4861
                     /note="3 copies 90 mer 71% conserved"
     repeat_region   4619. .4800
                     /note="7 copies 26 mer 79% conserved"
     repeat_region   8042. .8542
                     /note="LTR45 repeat: matches 1. .513 of consensus"
     repeat_region   8945. .9067
                     /note="MER5B repeat: matches 53. .177 of consensus"
     repeat_region   10365. .10706
                     /note="L2 repeat: matches 2350. .2674 of consensus"
     repeat_region   10990. .11142
                     /note="MLT1G repeat: matches 121. .283 of consensus"
     repeat_region   11228. .11322
                     /note="MLT1E repeat: matches 468. .566 of consensus"
     repeat_region   12021. .12363
                     /note="MER46B repeat: matches 1. .236 of consensus"
     repeat_region   13655. .13678
                     /note="12 copies 2 mer ag 95% conserved"
     repeat_region   13942. .14438
                     /note="L1MB5 repeat: matches 5679. .6174 of consensus"
     repeat_region   15351. .15666
                     /note="2 copies 158 mer 88% conserved"
     misc_feature    complement(15754. .16330)
                     /note="match: GSS: Em:AQ394110"
     repeat_region   15772. .16057
                     /note="MLT1E repeat: matches 267. .567 of consensus"
     repeat_region   16283. .16653
                     /note="MLT1A1 repeat: matches 1. .352 of consensus"
     repeat_region   16889. .17056
                     /note="MIR repeat: matches 59. .229 of consensus"
     repeat_region   17607. .17725
                     /note="AluSg/x repeat: matches 186. .302 of consensus"
     repeat_region   18349. .18558
                     /note="21 copies 10 mer atacacacat 68% conserved"
     repeat_region   18353. .18552
```

```
                     /note="100 copies 2 mer at 62% conserved"
     repeat_region   18357. .18564
                     /note="8 copies 26 mer 69% conserved"
     repeat_region   18358. .18567
                     /note="7 copies 30 mer 68% conserved"
     misc_feature    complement(18438. .18827)
                     /note="match: GSS: Em:AQ581588"
     misc_feature    complement(18532. .18952)
                     /note="match: GSS: Em:AQ633379"
     repeat_region   18813. .18944
                     /note="FLAM_A repeat: matches 1. .133 of consensus"
     misc_feature    complement(19334. .19731)
                     /note="match: GSS: Em:AQ122529"
     repeat_region   19472. .19617
                     /note="MIR repeat: matches 78. .218 of consensus"
     repeat_region   20285. .20513
                     /note="MLT1D repeat: matches 320. .501 of consensus"
     repeat_region   20695. .20832
                     /note="MIR repeat: matches 53. .189 of consensus"
     repeat_region   20924. .21387
                     /note="MLT1D repeat: matches 2. .505 of consensus"
     repeat_region   21818. .21867
                     /note="MER91A repeat: matches 1. .48 of consensus"
     repeat_region   22143. .22304
                     /note="MIR repeat: matches 31. .192 of consensus"
     repeat_region   22339. .22584
                     /note="MLT1A2 repeat: matches 1. .225 of consensus"
     repeat_region   22585. .23315
                     /note="LTR39 repeat: matches 1. .794 of consensus"
     repeat_region   23316. .23455
                     /note="MLT1A2 repeat: matches 225. .370 of consensus"
     repeat_region   24413. .24566
                     /note="MIR repeat: matches 29. .200 of consensus"
     misc_feature    24745. .25217
                     /note="match: GSS: Em:AQ403530"
     misc_feature    24758. .25162
                     /note="match: GSS: Em:AQ435639"
     repeat_region   25811. .25956
                     /note="MIR repeat: matches 54. .232 of consensus"
     repeat_region   27482. .27631
                     /note="5 copies 30 mer 77% conserved"
     repeat_region   27497. .27631
                     /note="9 copies 15 mer 79% conserved"
     misc_feature    27635
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 320bp by HindIII, SacI and SspI
                     restriction enzyme digest data."
     repeat_region   28011. .28380
                     /note="MSTB repeat: matches 1. .426 of consensus"
     repeat_region   29147. .29602
                     /note="LTR24 repeat: matches 1. .490 of consensus"
     repeat_region   30051. .30224
                     /note="2 copies 87 mer 92% conserved"
     repeat_region   32247. .32977
                     /note="L1PA16 repeat: matches 5423. .6154 of consensus"
     repeat_region   33453. .33754
                     /note="LTR8 repeat: matches 390. .691 of consensus"
     repeat_region   34281. .34449
                     /note="BUR1 repeat: matches 8180. .8350 of consensus"
     repeat_region   34808. .34911
                     /note="4 copies 26 mer 78% conserved"
     repeat_region   34811. .34902
                     /note="46 copies 2 mer tg 78% conserved"
     repeat_region   34813. .34902
                     /note="3 copies 30 mer 84% conserved"
     repeat_region   34821. .34910
                     /note="9 copies 10 mer tgtgtatgta 77% conserved"
     repeat_region   34914. .34991
                     /note="3 copies 26 mer 79% conserved"
     repeat_region   34916. .34975
                     /note="6 copies 10 mer tgtgtgtgta 78% conserved"
     repeat_region   35201. .35320
                     /note="MER94 repeat: matches 1. .123 of consensus"
     repeat_region   35497. .35835
                     /note="L1PA8 repeat: matches 5819. .6162 of consensus"
     repeat_region   35836. .36095
                     /note="AluY repeat: matches 39. .304 of consensus"
     repeat_region   36096. .40074
                     /note="L1PA8 repeat: matches 1897. .5819 of consensus"
     repeat_region   40075. .40337

  Query Match             68.0%;  Score 17;  DB 90;  Length 115551;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db   72877 AAGTCACAGAATTATTT 72861


RESULT  21
AC083945/c
LOCUS       AC083945   147706 bp    DNA             PLN       09-MAR-2001
DEFINITION  Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, complete sequence.
ACCESSION   AC083945
VERSION     AC083945.3  GI:12061428
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 147706)
  AUTHORS   Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., Bal,H.P., O'Shaughnessy,A.,
            Dedhia,N.N. and McCombie,W.R.
  TITLE     Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, Complete Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147706)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   3  (bases 1 to 147706)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   4  (bases 1 to 147706)
  AUTHORS   Bal,H.P., Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., O'Shaughnessy,A., Dedhia,N.N.
            and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1, Bungtown Road, Cold Spring Harbor, NY
            11743, USA
  REMARK    Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, Complete Sequence.
REFERENCE   5  (bases 1 to 147706)
  AUTHORS   Bal,H.P., Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., O'Shaughnessy,A., Dedhia,N.N.
            and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1, Bungtown Road, Cold Spring Harbor, NY
            11743, USA
```

```
  REMARK     Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
             Clone OSJNBa0058E19, Complete Sequence.
COMMENT      On Jan 10, 2001 this sequence version replaced gi:11038452.
             This sequence was finished as follows unless otherwise noted:  all
             regions were either double-stranded or sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by at least
             one plasmid subclone or more than one M13 subclone; and the
             assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .147706
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="X"
                     /clone="OSJNBa0058E19"
                     /clone_lib="HindIII"
     gene            58. .5309
                     /gene="OSJNBa0058E19.1"
                     /note="Histidine kinase-like protein"
     CDS             join(58. .156,495. .1156,1654. .2157,2536. .2693,2853. .2957,
                     4891. .5136,5200. .5309)
                     /gene="OSJNBa0058E19.1"
                     /codon_start=1
                     /product="Histidine-kinase like protein"
                     /protein_id="AAK13126.1"
                     /db_xref="GI:13129468"
                     /translation="MAVLVSDQVPDVLIGDPWRFRQIITNLVGNSMKFTEQGHIFIRV
                     HLIEEVKRKMEALDDTSPENIEVTANSKNTMPYNTLSGLEVANNRKTLESFRMFKDSS
                     DAIDSVNLLVTVEDTGIGITKDAQTRIFTPFMQADGSTSRTYGGTGIGLSITKRLVEL
                     MGGEIGFVSKPGVSSTFSFTAIFKENRKDPGDIKRYCPEPTPPDFQGMRALVVDGRCA
                     RAEVTMYHLRRLGIQCDLAATSESALSALLESCNSSVKSSLNMVLVDKEAWGEDSGLA
                     FFRCLIDLRLKGTLKSWQTMPKFFLLAGSITPADSDCLRLAGYSNSIRKPLRLSTVAA
                     CLSKALGVGLTGRRSRDNSLVLRSVLTGKNILVVDDNAVNRIVAAGALKKYGAIVTCV
                     DSGKEAISRLQPPHKFDACFMDVQMPEMDGFEATRLVRSVESKINDTIQAGEVSSEIY
                     GNKAHWHVPILAMTADVIQATFEGSSAKNPFGDNNIAKQKSYAVNREQKFDWTAIYVS
                     FEPTHPDSNTFNTGRAGPDRGRAVQPHWAPQIVGAPPHPPRSLYAAVAQVTGANMLCN
                     DTEQKLQTQRRSQFLPMRRDCRRTALHDSSRVVVVVGGGGGTSVRGAVTPWSSAIDCN
                     RIC"
     repeat_region   4968. .7539
                     /note="Micropon4 Oryza sativa gene, repeat sequence
                     Micropon"
     gene            complement(7197. .10518)
                     /gene="OSJNBa0058E19.2"
                     /note="Putative protein kinase Xa21"
     CDS             complement(join(7197. .7319,7361. .7668,7886. .8184,
                     8255. .8972,9073. .9099,9272. .9425,10121. .10354,
                     10369. .10518))
                     /gene="OSJNBa0058E19.2"
                     /codon_start=1
                     /product="Putative protein kinase Xa21"
                     /protein_id="AAK13127.1"
                     /db_xref="GI:13129469"
                     /translation="MNKLSLALASYLILFLAAAATAATTRCTFEIVVKTDGRRNAGTD
                     ARVSLQGPDADGRQPGVVGPDGRRPRLLREGEPRPVQGRRRLHAVGALQHGAHLRRLW
                     EQARVVRELRDGDAARAGETPVDDAPVANGPKNFAGLWRRSHHQPTTTYWVGKSTPKV
                     PQLVIMIQKLPRTAKPDTAGKPTGGEEVDLSSGAAAAASAAGRRCTSPTSHRARARSG
                     CAPSSPGSGSSRAAVQLRRVSRGCALFVYRAAEGASSTPTPPEEQTTRAAAAAVASPL
                     PALNLTNRGLIGQMSPLGNLTFLKFLFVPANSFTGEIPQSFGNMHHLQIIYLSNNTLQ
                     RRMPNLANCSNLKVLWLNGNNLVGQIPADLPQRFQSLQLSVNSLTGPIPVSVAKITTL
                     KRFSCLYNNIDGSIPNDFAKLPGLVYLHLGAKIIGSGELPSNIGNSLPDLQKFQLGGN
                     FFYGHIPNSLTNASKLNLIDIVSINSFTGVVPRSIGKLTKLSWLNLELNKFHAHSQKD
                     LEFMNSLANCTELQMSSIYGGVGGELAPSAARYTKSAHPRENPVSASKPNGPPSSSPN
                     PAKKARSRSALAPGATLVTYTRLPAAAAAPEDGSAPAVRGELDGDGAAGDDDAAAAAT
                     HLAERTPAQRSREVREKGEKRGNRKVSTWHPDMWGPHGSHADSAAM"
     repeat_region   7923. .9998
                     /note="Retrosat2 Ty3-Gypsy_retroelement"
     gene            18596. .19460
                     /gene="OSJNBa0058E19.3"
                     /note="Hypothetical protein"
     CDS             join(18596. .18726,18794. .18896,18917. .18991,19260. .19460)
                     /gene="OSJNBa0058E19.3"
                     /codon_start=1
                     /product="Hypothetical protein"
                     /protein_id="AAK13128.1"
                     /db_xref="GI:13129470"
                     /translation="MNLIVVQIRKMKSLFLLHSISSKAAMGLWPSARRCRRQMVTPLG
                     GHRSSASGESEQRMFSGGCACRAIDWMYPKGCMHGTHRSSDEVRVGLDSDDDAEDVPS
                     ALYLLHSNRNRRRDLVAAVHCVRSGAPAGEVAFPPNHCMIEAEIRGDGTGIERRRWNT
                     REKETIAAQ"
     gene            complement(22790. .40376)
                     /gene="OSJNBa0058E19.4"
                     /note="Putative gag-pol polyprotein"
     CDS             complement(join(22790. .23208,24472. .24598,24918. .25000,
                     25046. .25158,26244. .26431,26874. .26955,27044. .27126,
                     28396. .28467,28767. .28900,29535. .29647,30186. .30251,
                     30344. .30686,31565. .31711,32362. .32410,33038. .33286,
                     34586. .34714,35757. .36060,36212. .36315,37016. .37109,
                     37580. .37812,37836. .37921,38120. .38188,39131. .39252,
                     39319. .39444,39499. .40376))
                     /gene="OSJNBa0058E19.4"
                     /codon_start=1
                     /product="Putative gag-pol polyprotein"
                     /protein_id="AAK13129.1"
                     /db_xref="GI:13129471"
                     /translation="MDSIMSNGTWEVVERPYGCKPVGCKWVFKKKLRPDGTIEKYKAR
                     LVAKGYTQKEGEDFFDTYSPVARLTTIRVLLALAASHGLLIHQMDVKTAFLNGELEEE
                     IYMDQPDGYVLESQEGMVCKLLKSLYGLKQAPKQWHEKFDNTLTSAGFVVNEADKCVY
                     YRCGGGEGVILCLYVDDILIFGTSLNVIEEVKDFLSKSFEMTDLRVADVILNIKLLRG
                     DEGGITLVQSHYVDKVLSRFGYSDCKAAPTPYDPSVLLRKNRRIARDQLRYSQIIGSL
                     MYLASATRPDISFAVSKVMRYLKGTMSYGIHYAGYPKVLEGYSDSNWISDADEIKATS
                     GSTMEAELTALDTATVEAEWLRELLMDLPVVEKPVPAILMNCGARRGEARQTAATAAL
                     SRFVTDATRRTEEVLHDSCTSSSSSSVPELQNPSLVYGTSRCLRNSVRLRTCTGRRAI
                     KFLGSASSATAPVLRDYYYCIDAFVNMSMEIGDKEVNDINGGNTASKSSGGTFLGLNF
                     REILLSGRKFTWANSMPSPTYKKLDRVLEFTEEEVKHVVFQMEHKKAPGLDGFPAELI
                     EHKGNFKVFTKVMANRIALVAQKVIRPSQTAFFKGRNIMEGAIIHHETLHEMCKKKSD
                     GVILKLYFEKSYDKENWNFLQQALRMKSFYIWFKWIDQIVSGGSVAVKPKVCIGPDPV
                     LGPGLTLVLVCSTLLIKRGVETYKIYILRERAKPAVACSAASASARAPPSLPRVQPSA
                     AAAPVTAVAPEGHRQVRKVATVLIPSSVSAKDCRSAAAIDDPKTAAASTSLPSSSKPP
                     SFSLLCILACRFRRFEGGDYDPGKVRCFELSYVRGLVTTTQMVTCSAWAHDDMVTYCG
                     VVSCGYSGSPTRLESKRCGQDASVESLRTGGALDTRRTCPAPPNLAKQVEKHPWRYFR
                     QVGSHPAHHAPRGEHTSEDPTISDGAQEPGVIRITRLKGFLQDCVWEPASSDKTADKT
                     RLRVDCLKWPPLYLPDLGRTMNTNIGVSALNRSMTGRMTDGTQDVRRTIACVRRRAAR
                     HIPSNNDGSFERWRDKRRTANAGEGAASSGVPRHPRQGARPPASSPDKAGGQRLVLVP
                     LAGYGARPGSTNRPFKLQEISSRPKVWAKLPSPQRLGNREESGKLRLASTQTTWERFW
                     SGLTPGATVGNMGPGVPTYPHLWRPERSRYAAFNAVCSALNRSMTGRMTDGTQDVRRT
                     IACVRRRAARHIPSNNDGERRSIVGRNLEAWYELVSKVVNTALTDQKDRFFWKKNKDG
                     VVPRKNFIWKLKIPLKIKSFLWYLKERVIFMGTLWARTWPLLLKGVDEEMVKSYCKLL
                     EKRALEFFSMYVVTTMHPSIEEAFHRAPSMHASRIVVTTYKQHQEAEGCEKVKEEEPS
                     MQCGCRRTKFASKRLLSIQSSQRRKWLGFLVWPTTRSCEDHEWSAEVVAWGRCFAAVV
                     LGRAGEPPVGEWMMAVMAPKRIDAYCSIGVGRTG"
     repeat_region   25681. .29265
                     /note="retro02 retroelement"
     repeat_region   30888. .31407
                     /note="RIRE4 Oryza sativa gypsy-type retrotransposon"
     repeat_region   complement(30963. .32572)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     repeat_region   31529. .31692
                     /note="Truncator transposon from Oryza sativa"
     repeat_region   complement(31804. .32083)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     repeat_region   32562. .34237
                     /note="Truncator transposon"
     repeat_region   complement(32638. .33296)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     gene            complement(34475. .43713)
                     /gene="OSJNBa0058E19.5"
                     /note="Putative gag-pol polyprotein"
     CDS             complement(join(34475. .34662,35774. .35971,37596. .37812,
                     39105. .39266,39507. .40550,40983. .42461,42516. .42653,
                     42714. .43024,43070. .43269,43327. .43469,43702. .43713))
                     /gene="OSJNBa0058E19.5"
                     /codon_start=1
                     /product="Putative gag-pol polyprotein"
                     /protein_id="AAK13130.1"
                     /db_xref="GI:13129472"
```

```
  Query Match             68.0%;  Score 17;  DB 12;  Length 147706;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattatt 24
           |||||||||||||||||
Db   10686 CAAGTCACAGAATTATT 10670


RESULT  22
AC021110
LOCUS       AC021110   155172 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC021110
VERSION     AC021110.3  GI:8439978
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 10, 2000 this sequence version replaced gi:7021826.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0238P13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 73%
            Sequencing vector: plasmid; 27%
            Chemistry: Dye-primer ET; 73% of reads
            Chemistry: Dye-terminator Big Dye; 27% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151162 bases at least Q40
            Consensus quality: 152443 bases at least Q30
            Consensus quality: 153249 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 154672; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.69 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9265: contig of 9265 bp in length
            *    9266     9365: gap of unknown length
            *    9366    26504: contig of 17139 bp in length
            *   26505    26604: gap of unknown length
            *   26605    44300: contig of 17696 bp in length
            *   44301    44400: gap of unknown length
            *   44401    62250: contig of 17850 bp in length
            *   62251    62350: gap of unknown length
            *   62351   105744: contig of 43394 bp in length
            *  105745   105844: gap of unknown length
            *  105845   155172: contig of 49328 bp in length.

FEATURES             Location/Qualifiers
     source          1. .155172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-238P13"
     misc_feature    1. .9265
                     /note="assembly_name:Contig1"
     misc_feature    9366. .26504
                     /note="assembly_name:Contig2"
     misc_feature    26605. .44300
                     /note="assembly_name:Contig3
                     clone_end:T7
                     vector_side:right"
     misc_feature    44401. .62250
                     /note="assembly_name:Contig4"
     misc_feature    62351. .105744
                     /note="assembly_name:Contig5"
     misc_feature    105845. .155172
                     /note="assembly_name:Contig6"
BASE COUNT    51543 a  28385 c  28043 g  46698 t    503 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 66;  Length 155172;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db 111929 GCAAGTCACAGAATTAT 111945


RESULT  23
AP001390/c
LOCUS       AP001390   174815 bp    DNA             HTG       30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
            DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION   AP001390
VERSION     AP001390.2  GI:8117301
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-635B11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 174,815 genomic DNA of 18q23
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7259325.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-635B11
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
```

```
    Consensus quality: 159194 bases at least Q40
    Consensus quality: 167087 bases at least Q30
    Consensus quality: 170693 bases at least Q20
    Insert size: 172815; sum-of-contigs
    Quality coverage: 4.94x in Q20 bases; sum-of-contigs
--------------.
NOTE: This is a 'working draft' sequence. It currently consists of
21 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
         1     58823 contig of  58823 bp in length
     58924     73243 contig of  14320 bp in length
     73344     83572 contig of  10229 bp in length
     83673     94547 contig of  10875 bp in length
     94648    104597 contig of   9950 bp in length
    104698    114347 contig of   9650 bp in length
    114448    122080 contig of   7633 bp in length
    122181    128465 contig of   6285 bp in length
    128566    135791 contig of   7226 bp in length
    135892    142934 contig of   7043 bp in length
    143035    147999 contig of   4965 bp in length
    148100    151734 contig of   3635 bp in length
    151835    155576 contig of   3742 bp in length
    155677    159193 contig of   3517 bp in length
    159294    161359 contig of   2066 bp in length
    161460    163515 contig of   2056 bp in length
    163616    166614 contig of   2999 bp in length
    166715    169056 contig of   2342 bp in length
    169157    171527 contig of   2371 bp in length
    171628    173248 contig of   1621 bp in length
    173349    174815 contig of   1467 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     58823: contig of 58823 bp in length
*     58824 58923: gap of      100 bp
*     58924    73243: contig of 14320 bp in length
*     73244 73343: gap of      100 bp
*     73344    83572: contig of 10229 bp in length
*     83573 83672: gap of      100 bp
*     83673    94547: contig of 10875 bp in length
*     94548 94647: gap of      100 bp
*     94648   104597: contig of 9950 bp in length
*    104598 104697: gap of      100 bp
*    104698   114347: contig of 9650 bp in length
*    114348 114447: gap of      100 bp
*    114448   122080: contig of 7633 bp in length
*    122081 122180: gap of      100 bp
*    122181   128465: contig of 6285 bp in length
*    128466 128565: gap of      100 bp
*    128566   135791: contig of 7226 bp in length
*    135792 135891: gap of      100 bp
*    135892   142934: contig of 7043 bp in length
*    142935 143034: gap of      100 bp
*    143035   147999: contig of 4965 bp in length
*    148000 148099: gap of      100 bp
*    148100   151734: contig of 3635 bp in length
*    151735 151834: gap of      100 bp
*    151835   155576: contig of 3742 bp in length
*    155577 155676: gap of      100 bp
*    155677   159193: contig of 3517 bp in length
*    159194 159293: gap of      100 bp
*    159294   161359: contig of 2066 bp in length
*    161360 161459: gap of      100 bp
*    161460   163515: contig of 2056 bp in length
*    163516 163615: gap of      100 bp
*    163616   166614: contig of 2999 bp in length
*    166615 166714: gap of      100 bp
*    166715   169056: contig of 2342 bp in length
*    169057 169156: gap of      100 bp
*    169157   171527: contig of 2371 bp in length
*    171528 171627: gap of      100 bp
*    171628   173248: contig of 1621 bp in length
*    173249 173348: gap of      100 bp
*    173349   174815: contig of 1467 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q23"
                     /clone="RP11-635B11"
     misc_feature    1. .58823
                     /note="assembly_fragment"
     misc_feature    58924. .73243
                     /note="assembly_fragment"
     misc_feature    73344. .83572
                     /note="assembly_fragment"
     misc_feature    83673. .94547
                     /note="assembly_fragment"
     misc_feature    94648. .104597
                     /note="assembly_fragment"
     misc_feature    104698. .114347
                     /note="assembly_fragment"
     misc_feature    114448. .122080
                     /note="assembly_fragment"
     misc_feature    122181. .128465
                     /note="assembly_fragment"
     misc_feature    128566. .135791
                     /note="assembly_fragment"
     misc_feature    135892. .142934
                     /note="assembly_fragment"
     misc_feature    143035. .147999
                     /note="assembly_fragment"
     misc_feature    148100. .151734
                     /note="assembly_fragment"
     misc_feature    151835. .155576
                     /note="assembly_fragment"
     misc_feature    155677. .159193
                     /note="assembly_fragment"
     misc_feature    159294. .161359
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    161460. .163515
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    163616. .166614
                     /note="assembly_fragment"
     misc_feature    166715. .169056
                     /note="assembly_fragment"
     misc_feature    169157. .171527
                     /note="assembly_fragment"
     misc_feature    171628. .173248
                     /note="assembly_fragment"
     misc_feature    173349. .174815
                     /note="assembly_fragment"
BASE COUNT    52310 a  30923 c  32183 g  57399 t   2000 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 82;  Length 174815;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db   44384 GCAAGTCACAGAATTAT 44368
```

```
RESULT  24
AC002523/c
LOCUS       AC002523   228516 bp    DNA             PRI       05-MAY-1998
DEFINITION  Homo sapiens Xq28 BACs 360 F12, GSHB-555C13, complete sequence.
ACCESSION   AC002523 AC002374 AC002373
VERSION     AC002523.1  GI:2815518
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 228516)
  AUTHORS   Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
            Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,
            Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,
            Worley,K., Chen,E., Forcum,J., Arenson,A.D., Gorrell,J.H.,
            Brundage,E., Di,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 228516)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 228516)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 228516)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.

            Begining of sequence overlaps with AF007262, end of sequence
            overlaps with AF011889.
FEATURES             Location/Qualifiers
     source          1. .228516
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="360 F12, GSHB-555C13"
                     /chromosome="X"
     misc_feature    1. .149
                     /note="Positions 104645 to 104793 in AF007262 overlap with
                     postions 1-149 in this clone"
                     /note="Region: Overlap with AF007262"
     repeat_region   complement(487. .547)
                     /rpt_family="MER5B"
     repeat_region   complement(1270. .1577)
                     /rpt_family="MLT1E"
     repeat_region   complement(1653. .1708)
                     /rpt_family="MLT1D"
     repeat_region   complement(2282. .2306)
                     /rpt_family="AT_rich"
     repeat_region   3316. .3617
                     /rpt_family="MLT1B"
     repeat_region   3618. .3651
                     /rpt_family="(CA)n"
     repeat_region   complement(3658. .3944)
                     /rpt_family="AluSx"
     repeat_region   3949. .4012
                     /rpt_family="MLT1B"
     repeat_region   complement(4098. .4386)
                     /rpt_family="LINE2"
     repeat_region   4723. .4784
                     /rpt_family="MIR"
     repeat_region   5630. .5727
                     /rpt_family="(GA)n"
     repeat_region   7337. .7584
                     /rpt_family="LINE2"
     repeat_region   7678. .7709
                     /rpt_family="(CA)n"
     repeat_region   7710. .7863
                     /rpt_family="LINE2"
     repeat_region   complement(7981. .8151)
                     /rpt_family="LINE2"
     repeat_region   complement(8177. .8274)
                     /rpt_family="LINE2"
     repeat_region   complement(10118. .10192)
                     /rpt_family="LINE2"
     repeat_region   complement(10400. .10443)
                     /rpt_family="AT_rich"
     repeat_region   complement(10668. .10695)
                     /rpt_family="AT_rich"
     repeat_region   11811. .11896
                     /rpt_family="LINE2"
     repeat_region   12526. .12605
                     /rpt_family="LINE2"
     repeat_region   14892. .14932
                     /rpt_family="MLT1A2"
     repeat_region   15361. .15661
                     /rpt_family="AluSp"
     repeat_region   16706. .16765
                     /rpt_family="AT_rich"
     repeat_region   17193. .17764
                     /rpt_family="L1PA7"
     repeat_region   17811. .18912
                     /rpt_family="L1PA7"
     repeat_region   complement(18968. .19018)
                     /rpt_family="L1"
     repeat_region   19034. .19085
                     /rpt_family="L1"
     repeat_region   complement(19084. .19727)
                     /rpt_family="L1"
     repeat_region   19737. .19971
                     /rpt_family="L1"
     repeat_region   19971. .20406
                     /rpt_family="L1ME3A"
     repeat_region   20452. .20500
                     /rpt_family="AT_rich"
     repeat_region   20899. .21008
                     /rpt_family="(CAAAA)n"
     repeat_region   21061. .21098
                     /rpt_family="(TA)n"
     repeat_region   complement(21283. .21343)
                     /rpt_family="AT_rich"
     repeat_region   complement(21527. .22106)
                     /rpt_family="MLT1F"
     repeat_region   complement(22286. .22326)
                     /rpt_family="AT_rich"
     repeat_region   23171. .23316
                     /rpt_family="MER44A"
     repeat_region   23311. .23626
                     /rpt_family="MER44C"
     repeat_region   complement(23681. .23757)
                     /rpt_family="AT_rich"
     repeat_region   24613. .24924
                     /rpt_family="AluSx"
     repeat_region   complement(24927. .25278)
                     /rpt_family="MSTA"
```

```
     repeat_region   25554. .25668
                     /rpt_family="MIR"
     repeat_region   complement(27230. .27359)
                     /rpt_family="MER45"
     repeat_region   complement(28025. .28081)
                     /rpt_family="MER45"
     repeat_region   28227. .28312
                     /rpt_family="MIR"
     repeat_region   complement(28521. .28607)
                     /rpt_family="MIR"
     repeat_region   complement(29882. .29953)
                     /rpt_family="AT_rich"
     repeat_region   29960. .30154
                     /rpt_family="MIR"
     repeat_region   complement(30336. .30383)
                     /rpt_family="AT_rich"
     repeat_region   complement(31519. .31757)
                     /rpt_family="MIR"
     repeat_region   32198. .32905
                     /rpt_family="L1MA9"
     repeat_region   33664. .33841
                     /rpt_family="MIR"
     repeat_region   complement(34528. .34559)
                     /rpt_family="AT_rich"
     repeat_region   complement(34576. .34659)
                     /rpt_family="MIR"
     repeat_region   complement(34735. .35097)
                     /rpt_family="MSTD"
     repeat_region   complement(35249. .35546)
                     /rpt_family="AluY"
     repeat_region   complement(36693. .36767)
                     /rpt_family="(CA)n"
     repeat_region   complement(37299. .37786)
                     /rpt_family="MLT1D"
     STS             38010. .38096
                     /note="L47526, HUMSWX2596, Chr. X"
                     /db_xref="dbSTS: 17731"
     repeat_region   38232. .38255
                     /rpt_family="AT_rich"
     repeat_region   39138. .39218
                     /rpt_family="MIR"
     repeat_region   complement(39901. .40077)
                     /rpt_family="L1MC3"
     repeat_region   complement(40079. .40208)
                     /rpt_family="AluSg/x"
     repeat_region   complement(40209. .40353)
                     /rpt_family="L1MC3"
     repeat_region   complement(40354. .40646)
                     /rpt_family="MER47B"
     repeat_region   complement(40647. .40798)
                     /rpt_family="L1MC3"
     repeat_region   complement(40793. .42299)
                     /rpt_family="L1MC3"
     repeat_region   complement(42297. .42391)
                     /rpt_family="L1MD3"
     repeat_region   complement(42406. .45376)

  Query Match             68.0%;  Score 17;  DB 85;  Length 228516;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db   55846 AAGTCACAGAATTATTT 55830


RESULT  25
AP001690/c
LOCUS       AP001690    340000 bp    DNA             PRI       30-MAY-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 34/105.
ACCESSION   AP001690 AL163235 BA000005
VERSION     AP001690.1  GI:7768701
KEYWORDS    .
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
  TITLE     The DNA sequence of human chromosome 21. The chromosome 21 mapping
            and sequencing consortium
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
            Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
            Sciences Center, Human Genome Research Group * Institute of
            Molecular Biotechnology, Genome Analysis * Keio University School
            of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
            Analysis * Max-Planck Institute for Molecular Genetics (addresses
            see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717288.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163235: Submitted (10-Apr-2000).
FEATURES              Location/Qualifiers
     source           1. .340000
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
     source           <1. .114687
                      /note="Accession No. AP000469"
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="21"
```

```
                     /map="21q21.2"
                     /clone="R132H24, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
     repeat_region   complement(1162. .1526)
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   1527. .1893
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   1905. .2287
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   2386. .2549
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(4287. .4410)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(4489. .4670)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(6498. .6745)
                     /note="L1MA9"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   7600. .7882
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(7912. .8041)
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(8745. .9054)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(9504. .9794)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   9983. .10564
                     /note="L1MB8"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   10720. .10787
                     /note="(TAGA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   11137. .11436
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     STS             13135. .13345
                     /standard_name="D21S1914"
                     /note="AFM344wf5
                     Accession No. Z51475"
     repeat_region   13259. .13313
                     /note="(TG)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(14329. .14632)
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   14625. .14852
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   14853. .15154
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   15155. .15634
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(15694. .15808)
                     /note="L1MA6"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   15840. .16145
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   17627. .18220
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   18201. .18826
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   18827. .19126
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   19127. .19352
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   19353. .20555
                     /note="L1PA7"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   20557. .21024
                     /note="L1P5"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   21026. .21055
                     /note="(TG)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(21414. .21879)
                     /note="L1MA5A"
                     /rpt_family="LINE/L1"

  Query Match             68.0%;  Score 17;  DB 91;  Length 340000;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db 141812 AAGTCACAGAATTATTT 141796


RESULT  26
HSAPAF20/c
LOCUS       HSAPAF20      605 bp    DNA             PRI       03-AUG-2000
DEFINITION  Homo sapiens APAF1 gene, exon 12, 3' end.
ACCESSION   AF098889
VERSION     AF098889.1  GI:4191520
KEYWORDS    .
SEGMENT     20 of 47
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Bala,S., Oliver,H., Renault,B., Montgomery,K., Dutta,S., Rao,P.,
```

```
            Houldsworth,J., Kucherlapati,R., Wang,X., Chaganti,R.S. and
            Murty,V.V.
  TITLE     Genetic analysis of the APAF1 gene in male germ cell tumors
  JOURNAL   Genes Chromosomes Cancer 28 (3), 258-268 (2000)
  MEDLINE   20320642
   PUBMED   10862031
REFERENCE   2  (bases 1 to 605)
  AUTHORS   Bala,S., Oliver,H., Renault,B., Dutta,S., Montgomery,K.,
            Kucherlapati,R., Wang,X., Chaganti,R.S.K. and Murty,V.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1998) Pathology, College of Physicians & Surgeons
            of Columbia University, 630 West 168th Street, New York, New York
            10032, USA
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     exon            <1. .70
                     /gene="APAF1"
                     /number=12
     intron          71. .>605
                     /gene="APAF1"
                     /number=12
BASE COUNT      184 a     93 c    134 g    192 t      2 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 93;  Length 605;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattatt 24
          ||||||||||||||||
Db    425 AAGTCACAGAATTATT 410


RESULT  27
G07181/c
LOCUS       G07181       1829 bp    DNA             STS       19-OCT-1995
DEFINITION  human STS WI-9211.
ACCESSION   G07181
VERSION     G07181.1  GI:860426
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1829)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: CAGCCACACTAACTGCATTAGC
            Primer B: GACAGAAATGGATTGTATAAATGGC
            STS size: 111
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Prepared with primer pairs derived from D13644 -- Unigene.
FEATURES             Location/Qualifiers
     source          1. .1829
                     /organism="Homo sapiens"
     STS             891. .1001
     primer_bind     891. .912
     primer_bind     complement(977. .1001)
BASE COUNT      539 a    321 c    320 g    590 t     59 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 54;  Length 1829;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db   1689 CAAGTCACAGAATTAT 1674


RESULT  28
D13644/c
LOCUS       D13644       4602 bp    mRNA            PRI       10-JUL-1997
DEFINITION  Human mRNA for KIAA0019 gene, complete cds.
ACCESSION   D13644
VERSION     D13644.1  GI:1531551
KEYWORDS    KIAA0019.
SOURCE      Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4602)
  AUTHORS   Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1992) to the DDBJ/EMBL/GenBank databases. Nobuo
            Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
            Yana, Kisarazu, Chiba 292,, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://www.kazusa.or.jp, Tel:0438-52-3930, Fax:0438-52-3931)
REFERENCE   2  (bases 1 to 4602)
  AUTHORS   Nomura,N., Miyajima,N., Kawarabayasi,Y. and Tabata,S.
  TITLE     Prediction of new human genes by entire sequencing of randomly
            sampled cDNA clones
  JOURNAL   Unpublished (1994)
REFERENCE   3  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1
  JOURNAL   DNA Res. 1 (1), 27-35 (1994)
  MEDLINE   96051387
  REMARK    Erratum:[[published erratum appears in DNA Res 1995 Aug
            31;2(4):211]]
REFERENCE   4  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
```

```
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1 (supplement)
  JOURNAL   DNA Res. 1 (1), 47-56 (1994)
  MEDLINE   96051389
REFERENCE   5  (bases 1 to 4602)
  AUTHORS   Matoskova,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N.,
            Robbins,K.C. and Di Fiore,P.P.
  TITLE     RN-tre identifies a family of tre-related proteins displaying a
            novel potential protein binding domain
  JOURNAL   Oncogene 12 (12), 2563-2571 (1996)
  MEDLINE   96293402
COMMENT     On Sep 10, 1996 this sequence version replaced gi:286016.
FEATURES             Location/Qualifiers
     source          1. .4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="KG-1"
                     /cell_type="myoblast"
                     /chromosome="10"
                     /map="10p13"
                     /sex="male"
     5'UTR           <1. .279
     gene            280. .2766
                     /gene="KIAA0019"
     CDS             280. .2766
                     /gene="KIAA0019"
                     /codon_start=1
                     /product="protein related N-ternimus of tre oncogene"
                     /protein_id="BAA02807.1"
                     /db_xref="GI:2104571"
                     /translation="MNSDQDVALKLAQERAEIVAKYDRGREGAEIEPWEDADYLVYKV
                     TDRFGFLHEEELPDHNVAVERQKHLEIERTTKWLKMLKGWEKYKNTEKFHRRIYKGIP
                     LQLRGEVWALLLEIPKMKEETRDLYSKLKHRARGCSPDIRQIDLDVNRTFRDHIMFRD
                     RYGVKQQSLFHVLAAYSIYNTEVGYCQGMSQITALLLMYMNEEDAFWALVKLFSGPKH
                     AMHGFFVQGFPKLLRFQEHHEKILNKFLSKLKQHLDSQEIYTSFYTMKWFFQCFLDRT
                     PFTLNLRIWDIYIFEGERVLTAMSYTILKLHKKHLMKLSMEELVEFFQETLAKDFFFE
                     DDFVIEQLQISMTELKRAKLDLPEPGKEDEYPKKPLGQLPPELQSWGVHHLSNGQRSV
                     GRPSPLASGRRESGAPHRRHEHSPHPQSRTGTPERAQPPRRKSVEEESKKLKDEADFQ
                     RKLPSGPQDSSRQYNHAAANQNSNATSNIRKEFVPKWNKPSDVSATERTAKYTMEGKG
                     RAAHPALAVTVPGPAEVRVSNVRPKMKALDAEDGKRGSTASQYDNVPGPELDSGASVE
                     EALERAYSQSPRHALYPPSPRKHAEPSSSPSKVSNKFTFKVQPPSHARYPSQLDGEAR
                     GLAHPPSYSNPPVYHGNSPKHFPTANSSFASPQFSPGTQLNPSRRPHGSTLSVSASPE
                     KSYSRPSPLVLPSSRIEVLPVDTGAGGYSGNSGSPKNGKLIIPPVDYLPDNRTWSEVS
                     YTYRPETQGQSWTRDASRGNLPKYSAFQLAPFQDHGLPAVSVDSPVRYKASPAAEDAS
                     PSGYPYSGPPPPAYHYRNRDGLSIQESVLL"
     3'UTR           2767. .4602
BASE COUNT     1340 a   1041 c    999 g   1222 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 91;  Length 4602;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db    4462 CAAGTCACAGAATTAT 4447


RESULT  29
LLA012388
LOCUS       LLA012388    5469 bp    DNA             BCT       29-OCT-1999
DEFINITION  Lactococcus lactis dar gene and genes encoding five hypothetical
            proteins.
ACCESSION   AJ012388
VERSION     AJ012388.1  GI:6165401
KEYWORDS    acetoin(diacetyl)reductase; dar gene; ORF1; ORF3; ORF4; ORF5; ORF6;
            putative.
SOURCE      Lactococcus lactis.
  ORGANISM  Lactococcus lactis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 5469)
  AUTHORS   Aungpraphapornchai,P., Griffin,H.G. and Gasson,M.J.
  TITLE     Cloning, DNA sequence analysis, and deletion of a gene encoding
            diacetyl-acetoin reductase from Lactococcus lactis
  JOURNAL   DNA Seq. 10 (3), 163-172 (1999)
  MEDLINE   20112344
REFERENCE   2  (bases 1 to 5469)
  AUTHORS   Aungpraphapornchai,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1998) Aungpraphapornchai P., Genetics and
            Microbiology, Institute of Food Research, Colney lane, Norwich,
            Norfolk, NR4 7UA, UK
FEATURES             Location/Qualifiers
     source          1. .5469
                     /organism="Lactococcus lactis"
                     /sub_species="cremoris"
                     /strain="MG1363"
                     /db_xref="taxon:1358"
     RBS             77. .80
                     /note="ORF1"
     CDS             87. .950
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59825.1"
                     /db_xref="GI:6165402"
                     /translation="MNPKNRNIIIAVAVLILVALVAFFSLNHQGGVKASAGEKTVKVG
                     IMSGDKQDQEVWKSVANTAKEKYDLKLKFVYFSDYNQPNEALLSGDIDINAFQSYNYV
                     KTWNKAHKSDIVAVGNTYITPMHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQ
                     SAGLLKLTTSDSSKLVGLPDITENPHQLKFKEVDASQTPRALDSVALSVVNYNYATAA
                     SLPKSESVFMEPLNKTSAQYINFIATTSKEKNNKVYKEVAKAYASKATEKAIKEQYPD
                     GGELPAWDLKL"
     RBS             1005. .1009
                     /gene="dar"
     gene            1005. .1789
                     /gene="dar"
     CDS             1019. .1789
                     /gene="dar"
                     /codon_start=1
                     /transl_table=11
                     /product="diacetyl(acetoin) reductase"
                     /protein_id="CAB59826.1"
                     /db_xref="GI:6165403"
                     /translation="MGKLAIVTGSGQGIGEGIVYRLAKDGFAVAVADINPQTAKNVAK
                     TLIDKGFEAKAYVLDVANRSAVFQLIEQAVSDFGELAVFVNNAGVAFIDSIIDSDSER
                     VERLFDVNIKGTYWGIQAAATQFKKQGHGGRIINAASLASVEGSALQSAYSASKFAIR
                     GLTQSAAKELAKDYITVNAYNPGIVRTAMRDEIDKKTAQIKGISIEQQQQNCLNEIYL
                     GREGRPEDVAEVVAWFASDAAEYITGQSILVDGGMRFN"
     RBS             1794. .1798
                     /note="ORF3"
     CDS             1805. .2665
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59827.1"
                     /db_xref="GI:6165404"
                     /translation="MNVKNRRILITIIILVFIIIVGGIFAFSHSGNKSKVSSKIVKIG
                     LMPGGKQEDVIWKQVQKNAKDQFGITLKFVNFTDGDEPNKALVNHEVDLNAFQHYAYL
                     KSWNKANNGNIVSIGDTIITPIHLYSTKYKKVDEIPDKSTIAIPNDITNESRALYVLK
                     NAGLIKLDTSRGVLATVKDIRENPKSLIIKEIDASQTPRALDSVAAAVINYNFAISAK
                     NSDKESIYQEPLNEDSAQWINFIAANQSDKNNKVYKEVVKAYEQKNIADIIKKEYPDG
                     GELPAWNLKL"
     RBS             2776. .2781
                     /note="ORF4"
     CDS             2789. .3895
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59828.1"
                     /db_xref="GI:6165405"
                     /translation="MTAIIELNNLSVQFHQKGRLVTAVKNATLHIEKGDIYGVIGYSG
```

```
                     AGKSTLVRTINLLQKPTEGQIVINGEKIFDSENPVKFTGAKLREFRQKIGMIFQHFNL
                     LSEKTVFNNVAFALQHSQIEDKNGKKRYLTKKEKNDKVTELLKLVDLADLSDKYPAQL
                     SGGQKQRVAIARALANDPEILISDEGTSALDPKTTNQILDLLKSLHEKLGITVVLITH
                     EMQVVKEIANKVAVMQNGEIIEQNSLIDIFAQPKEALTKQFIETTSSVNRFIASLSKT
                     ELLAQLADDEELIHLDYSGSELEDPVVSDITKKFDVTTNIFYGNVELLQGQPFGSLVL
                     TLKGSSEHRAAAKAYFVERHLKFEVLGKIERTVD"
     RBS             3878. .3884
                     /note="ORF5"
     CDS             3895. .4590
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59829.1"
                     /db_xref="GI:6165406"
                     /translation="MAEWFAHTFPNVVYLGWTGETGWWTAIVQTLYMTFISALIGGLL
                     GLIFGIGVVVTAEDGITPNRPLFWILDKIVSIGRAFPFIILLAAIAPLTKILVGTQIG
                     VTAALVPLALGVAPFYARQVQASLESVDHGKVEAAQTVGADFLDIVFTVYLREELASL
                     IRVSTVTLISLIGLTAMAGAIGAGGLGNTAISYGYNRFANDVTWFATILILIFVLLVQ
                     LVGDFLARRVSHR"
     RBS             4621. .4627
                     /note="ORF6"
     CDS             4633. .5181
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59830.1"
                     /db_xref="GI:6165407"
                     /translation="MKNNSVKIVVATGIGAALFVIIGWLINIPTPIPNTSIQLQYAVL
                     ALFSALFGPLAGFLIGFIGHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQG
                     IFGNKEIIRFNIVQFLANVVVWGLIAPIGDILVYSEPANKVFTQGVVAGLVNALTIAV
                     AGTLLLKLYAATRTKSGTLDKE"
     terminator      5210. .5235
                     /note="putative"
BASE COUNT     1796 a    858 c   1134 g   1681 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 3;  Length 5469;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattatt 24
          ||||||||||||||||
Db   3204 AAGTCACAGAATTATT 3219


RESULT  30
AL161800/c
LOCUS       AL161800    43741 bp    DNA             PRI       04-APR-2001
DEFINITION  Human DNA sequence from clone RP11-318C17 on chromosome 20.
            Contains part of a novel gene and GSSs, complete sequence.
ACCESSION   AL161800
VERSION     AL161800.10  GI:8246908
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 43741)
  AUTHORS   Wilson,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 4, 2000 this sequence version replaced gi:8217673.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-318C17 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP5-974N19 is at 43642 in this sequence.
            The true right end of clone RP5-858M22 is at 100 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-318C17 is from
            the library RPCI-11.2 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .43741
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-318C17"
                     /clone_lib="RPCI-11.2"
     repeat_region   3. .293
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   390. .683
                     /note="AluSg repeat: matches 1. .286 of consensus"
     repeat_region   788. .3327
                     /note="L1PA16 repeat: matches 3587. .6157 of consensus"
     misc_feature    3411. .3820
                     /note="match: GSS: Em:AQ371040"
     repeat_region   4157. .4477
                     /note="L2 repeat: matches 2348. .2710 of consensus"
     misc_feature    complement(6523. .6967)
                     /note="match: GSS: Em:AQ236127"
     misc_feature    complement(6607. .6950)
                     /note="match: GSS: Em:AQ110431"
     repeat_region   6676. .6820
                     /note="MIR repeat: matches 79. .229 of consensus"
     repeat_region   7231. .7405
                     /note="MIR repeat: matches 42. .206 of consensus"
     repeat_region   7416. .7551
                     /note="L2 repeat: matches 2584. .2710 of consensus"
     repeat_region   9521. .9605
                     /note="MLT1C repeat: matches 433. .523 of consensus"
     mRNA            complement(join(9542. .9883,13385. .>13593))
                     /gene="bA318C17.1"
                     /note="continued from dJ974N19.2 in Em:AL138808
                     match: ESTs: Em:AL542749 Em:AL569592"
                     /evidence=not_experimental
                     /product="bA318C17.1 (novel transcript)"
     polyA_site      complement(9542)
                     /gene="bA318C17.1"
     gene            complement(join(9542. .9883,13385. .13593))
                     /gene="bA318C17.1"
     polyA_signal    complement(9565. .9570)
                     /gene="bA318C17.1"
     repeat_region   9805. .9884
                     /note="MLT1G repeat: matches 59. .137 of consensus"
     repeat_region   9885. .10320
                     /note="MER4B repeat: matches 1. .437 of consensus"
     repeat_region   10325. .10397
                     /note="MER4C repeat: matches 393. .465 of consensus"
     repeat_region   10682. .11085
```

```
                     /note="L2 repeat: matches 1981. .2406 of consensus"
     repeat_region   11167. .11657
                     /note="L2 repeat: matches 2263. .2750 of consensus"
     repeat_region   11872. .12002
                     /note="L2 repeat: matches 2621. .2750 of consensus"
     repeat_region   12681. .12976
                     /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region   14480. .15774
                     /note="L2 repeat: matches 1424. .2749 of consensus"
     repeat_region   15775. .16067
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   16068. .16216
                     /note="L2 repeat: matches 1285. .1424 of consensus"
     repeat_region   16713. .16865
                     /note="L2 repeat: matches 2545. .2699 of consensus"
     repeat_region   17484. .17764
                     /note="AluSx repeat: matches 6. .296 of consensus"
     repeat_region   17813. .18013
                     /note="L1MB2 repeat: matches 5966. .6164 of consensus"
     repeat_region   18014. .18267
                     /note="AluSc repeat: matches 1. .252 of consensus"
     repeat_region   18268. .18773
                     /note="L1MB2 repeat: matches 5450. .5966 of consensus"
     repeat_region   19067. .19374
                     /note="AluSg1 repeat: matches 1. .304 of consensus"
     repeat_region   19513. .19811
                     /note="AluYa8 repeat: matches 1. .305 of consensus"
     repeat_region   21304. .21613
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   22998. .23343
                     /note="MLT1A1 repeat: matches 1. .353 of consensus"
     repeat_region   23736. .23865
                     /note="AluSg repeat: matches 1. .130 of consensus"
     repeat_region   24570. .25025
                     /note="MLT1C repeat: matches 1. .466 of consensus"
     repeat_region   25235. .26166
                     /note="L1M4 repeat: matches 3073. .4010 of consensus"
     misc_feature    complement(26577. .27248)
                     /note="match: GSS: Em:AQ036701"
     repeat_region   28315. .28629
                     /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region   28933. .29096
                     /note="MLT1I repeat: matches 245. .409 of consensus"
     repeat_region   29893. .30110
                     /note="MIR repeat: matches 25. .245 of consensus"
     misc_feature    complement(30032. .30494)
                     /note="match: GSS: Em:AQ723457"
     repeat_region   31501. .31807
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   32165. .32267
                     /note="L2 repeat: matches 2554. .2646 of consensus"
     repeat_region   32268. .32752
                     /note="MER66B repeat: matches 1. .486 of consensus"
     repeat_region   32753. .33089
                     /note="L2 repeat: matches 2213. .2554 of consensus"
     repeat_region   33351. .33651
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   33662. .33966
                     /note="AluY repeat: matches 1. .295 of consensus"
     repeat_region   34730. .34818
                     /note="MIR repeat: matches 43. .145 of consensus"
     misc_feature    complement(34862. .35287)
                     /note="match: GSS: Em:AQ404224"
     misc_feature    complement(35239. .35778)
                     /note="match: GSS: Em:AQ634531"
     misc_feature    complement(35374. .35798)
                     /note="match: GSS: Em:AQ565906"
     misc_feature    36173. .36533
                     /note="match: GSS: Em:AQ231637"
     repeat_region   36392. .36553
                     /note="MER5A repeat: matches 4. .175 of consensus"
     repeat_region   36668. .36961
                     /note="AluSg repeat: matches 1. .297 of consensus"
     repeat_region   37033. .37258
                     /note="MIR repeat: matches 15. .252 of consensus"
     misc_feature    38080. .38507
                     /note="match: GSS: Em:B36377"
     misc_feature    38087. .38612
                     /note="match: GSS: Em:AQ680468"
     repeat_region   38349. .38418
                     /note="35 copies 2 mer tg 74% conserved"
     repeat_region   38694. .38979
                     /note="AluJo repeat: matches 1. .286 of consensus"
     misc_feature    40625. .41129
                     /note="match: GSS: Em:AQ680145"
     repeat_region   41704. .41819
                     /note="MIR repeat: matches 57. .177 of consensus"
     repeat_region   41842. .42034
                     /note="L1MC4 repeat: matches 7779. .7969 of consensus"
     repeat_region   42037. .42226
                     /note="MLT1J repeat: matches 45. .247 of consensus"
     misc_feature    complement(42848. .43207)
                     /note="match: GSS: Em:AQ385536"
     repeat_region   42966. .43258
                     /note="L2 repeat: matches 1138. .1493 of consensus"
     repeat_region   43259. .43555
                     /note="AluJb repeat: matches 10. .303 of consensus"
     repeat_region   43556. .43741
                     /note="L2 repeat: matches 970. .1138 of consensus"
BASE COUNT    12960 a   8117 c   8797 g  13867 t
ORIGIN


  Query Match           64.0%;  Score 16;  DB 90;  Length 43741;
  Best Local Similarity  100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   10623 AAGTCACAGAATTATT 10608


RESULT  31
AC074073/c
LOCUS       AC074073    68894 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-791I6 map 3, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC074073
VERSION     AC074073.1  GI:9170900
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68894)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-791I6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68894)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
```

```
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7127
                Center clone name: 791_I_6
            --------------.
            * NOTE: This record contains 83 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      747: contig of 747 bp in length
            *      748  847: gap of      100 bp
            *      848     1579: contig of 732 bp in length
            *     1580  1679: gap of      100 bp
            *     1680     2425: contig of 746 bp in length
            *     2426  2525: gap of      100 bp
            *     2526     3254: contig of 729 bp in length
            *     3255  3354: gap of      100 bp
            *     3355     4075: contig of 721 bp in length
            *     4076  4175: gap of      100 bp
            *     4176     4892: contig of 717 bp in length
            *     4893  4992: gap of      100 bp
            *     4993     5717: contig of 725 bp in length
            *     5718  5817: gap of      100 bp
            *     5818     6558: contig of 741 bp in length
            *     6559  6658: gap of      100 bp
            *     6659     7371: contig of 713 bp in length
            *     7372  7471: gap of      100 bp
            *     7472     8209: contig of 738 bp in length
            *     8210  8309: gap of      100 bp
            *     8310     9046: contig of 737 bp in length
            *     9047  9146: gap of      100 bp
            *     9147     9888: contig of 742 bp in length
            *     9889  9988: gap of      100 bp
            *     9989    10723: contig of 735 bp in length
            *    10724 10823: gap of      100 bp
            *    10824    11553: contig of 730 bp in length
            *    11554 11653: gap of      100 bp
            *    11654    12391: contig of 738 bp in length
            *    12392 12491: gap of      100 bp
            *    12492    13214: contig of 723 bp in length
            *    13215 13314: gap of      100 bp
            *    13315    14034: contig of 720 bp in length
            *    14035 14134: gap of      100 bp
            *    14135    14864: contig of 730 bp in length
            *    14865 14964: gap of      100 bp
            *    14965    15700: contig of 736 bp in length
            *    15701 15800: gap of      100 bp
            *    15801    16535: contig of 735 bp in length
            *    16536 16635: gap of      100 bp
            *    16636    17357: contig of 722 bp in length
            *    17358 17457: gap of      100 bp
            *    17458    18192: contig of 735 bp in length
            *    18193 18292: gap of      100 bp
            *    18293    19030: contig of 738 bp in length
            *    19031 19130: gap of      100 bp
            *    19131    19857: contig of 727 bp in length
            *    19858 19957: gap of      100 bp
            *    19958    20677: contig of 720 bp in length
            *    20678 20777: gap of      100 bp
            *    20778    21514: contig of 737 bp in length
            *    21515 21614: gap of      100 bp
            *    21615    22336: contig of 722 bp in length
            *    22337 22436: gap of      100 bp
            *    22437    23190: contig of 754 bp in length
            *    23191 23290: gap of      100 bp
            *    23291    24024: contig of 734 bp in length
            *    24025 24124: gap of      100 bp
            *    24125    24859: contig of 735 bp in length
            *    24860 24959: gap of      100 bp
            *    24960    25700: contig of 741 bp in length
            *    25701 25800: gap of      100 bp
            *    25801    26544: contig of 744 bp in length
            *    26545 26644: gap of      100 bp
            *    26645    27367: contig of 723 bp in length
            *    27368 27467: gap of      100 bp
            *    27468    28196: contig of 729 bp in length
            *    28197 28296: gap of      100 bp
            *    28297    29028: contig of 732 bp in length
            *    29029 29128: gap of      100 bp
            *    29129    29843: contig of 715 bp in length
            *    29844 29943: gap of      100 bp
            *    29944    30633: contig of 690 bp in length
            *    30634 30733: gap of      100 bp
            *    30734    31472: contig of 739 bp in length
            *    31473 31572: gap of      100 bp
            *    31573    32319: contig of 747 bp in length
            *    32320 32419: gap of      100 bp
            *    32420    33151: contig of 732 bp in length
            *    33152 33251: gap of      100 bp
            *    33252    33977: contig of 726 bp in length
            *    33978 34077: gap of      100 bp
            *    34078    34818: contig of 741 bp in length
            *    34819 34918: gap of      100 bp
            *    34919    35632: contig of 714 bp in length
            *    35633 35732: gap of      100 bp
            *    35733    36474: contig of 742 bp in length
            *    36475 36574: gap of      100 bp
            *    36575    37300: contig of 726 bp in length
            *    37301 37400: gap of      100 bp
            *    37401    38137: contig of 737 bp in length
            *    38138 38237: gap of      100 bp
            *    38238    38960: contig of 723 bp in length
            *    38961 39060: gap of      100 bp
            *    39061    39789: contig of 729 bp in length
            *    39790 39889: gap of      100 bp
            *    39890    40617: contig of 728 bp in length
            *    40618 40717: gap of      100 bp
            *    40718    41442: contig of 725 bp in length
            *    41443 41542: gap of      100 bp
            *    41543    42278: contig of 736 bp in length
            *    42279 42378: gap of      100 bp
            *    42379    43117: contig of 739 bp in length
            *    43118 43217: gap of      100 bp
            *    43218    43961: contig of 744 bp in length
            *    43962 44061: gap of      100 bp
            *    44062    44786: contig of 725 bp in length
            *    44787 44886: gap of      100 bp
            *    44887    45614: contig of 728 bp in length
            *    45615 45714: gap of      100 bp
            *    45715    46458: contig of 744 bp in length
            *    46459 46558: gap of      100 bp
            *    46559    47292: contig of 734 bp in length
            *    47293 47392: gap of      100 bp
            *    47393    48107: contig of 715 bp in length
            *    48108 48207: gap of      100 bp
```

```
            *    48208    48935: contig of 728 bp in length
            *    48936 49035: gap of      100 bp
            *    49036    49765: contig of 730 bp in length
            *    49766 49865: gap of      100 bp
            *    49866    50604: contig of 739 bp in length
            *    50605 50704: gap of      100 bp
            *    50705    51434: contig of 730 bp in length
            *    51435 51534: gap of      100 bp
            *    51535    52266: contig of 732 bp in length
            *    52267 52366: gap of      100 bp
            *    52367    53096: contig of 730 bp in length
            *    53097 53196: gap of      100 bp
            *    53197    53936: contig of 740 bp in length
            *    53937 54036: gap of      100 bp
            *    54037    54770: contig of 734 bp in length
            *    54771 54870: gap of      100 bp
            *    54871    55574: contig of 704 bp in length
            *    55575 55674: gap of      100 bp
            *    55675    56392: contig of 718 bp in length
            *    56393 56492: gap of      100 bp
            *    56493    57227: contig of 735 bp in length
            *    57228 57327: gap of      100 bp

  Query Match             64.0%;  Score 16;  DB 75;  Length 68894;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   58360 AGTCACAGAATTATTT 58345


RESULT  32
AC012679/c
LOCUS       AC012679   105937 bp    DNA             PLN       19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence,
            complete sequence.
ACCESSION   AC012679
VERSION     AC012679.5  GI:12324196
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 105937)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105937)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 105937)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280804.
            Address all correspondence to:at@tigr.org

            BAC clone F25P22 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .105937
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="nga111"
                     /clone="F25P22"
     mRNA            join(<1. .76,158. .299,430. .710,789. .950,1031. .1084,
                     1171. .1512)
                     /gene="F25P22.1"
     gene            1. .1512
                     /gene="F25P22.1"
                     /note="similar to
                     S-adenosyl-methionine-sterol-C-methyltransferase
                     GB:AAC34951 [Nicotiana tabacum]"
     CDS             join(<1. .76,158. .299,430. .710,789. .950,1031. .1084,
                     1171. .1344)
                     /gene="F25P22.1"
                     /codon_start=2
                     /product="putative
                     S-adenosyl-methionine-sterol-C-methyltransferase, 5'
                     partial; 1-1344"
                     /protein_id="AAG52075.1"
                     /db_xref="GI:12324209"
                     /translation="NSYELSLVSCKCIGAYVRNKKNQNQICWLWQKVSSDNDRGFQRF
                     LDNVQYKSSGILRYERVFGEGFVSTGGLETTKEFVDMLDLKPGQKVLDVGCGIGGGDF
                     YMAENFDVDVVGIDLSVNMISFALEHAIGLKCSVEFEVADCTKKEYPDNTFDVIYSRD
                     TILHIQDKPALFRRFYKWLKPGGKVLITDYCRSPKTPSPDFAIYIKKRGYDLHDVQAY
                     GQMLRDAGFEEVIAEDRTDQFMKVLKRELDAVEKEKEEFISDFSKEDYEDIIGGWKSK
                     LLRSSSGEQKWGLFIAKRN"
     mRNA            join(<6321. .6579,6690. .6820,6910. .7146,7259. .7502,
                     7588. .>7751)
                     /gene="F25P22.2"
     gene            6321. .7751
                     /gene="F25P22.2"
                     /note="contains Pfam profile: PF00657 lipase/acylhydrolase
                     with GDSL-like motif"
     CDS             join(6321. .6579,6690. .6820,6910. .7146,7259. .7502,
                     7588. .7751)
                     /gene="F25P22.2"
                     /codon_start=1
                     /product="putative lipase/acylhydrolase; 6321-7751"
                     /protein_id="AAG52082.1"
                     /db_xref="GI:12324216"
                     /translation="MNCLMFFKMLLAFSFISLFYVGNAQQSYGNSTVSALFAFGDSIL
                     DTGNNNLLLSVSKVNFYPYGRDFIGGRATGRFGNGRVFSDIIAEGLGLKNLLPAYRDP
                     YLWNNDLTTGVCFASGGSGLDPITARTTGSIWVSDQVTDFQNYITRLNGVVGNQEQAN
                     AVISNAVYLISAGNNDIAITYFTTGARRLQYTLPAYNDQLVSWTRDLIKSLYDMGARK
                     FAVMGTLPLGCLPGARALTRACELFVNQGAAMFNQQLSADIDNLGATFPGAKFVYVDM
                     YNPLLGLIINPQASGFIDVADACCCTPTHLIPCLDASRYVFWDVAHPTQKSYETIAPQ
                     IIENIKAKLA"
     mRNA            join(<9376. .9496,9643. .9967,10550. .>10898)
                     /gene="F25P22.3"
     gene            9376. .10898
                     /gene="F25P22.3"
                     /note="contains Pfam profile: PF00314 thaumatin family"
     CDS             join(9376. .9496,9643. .9967,10550. .10898)
                     /gene="F25P22.3"
```

```
                     /codon_start=1
                     /product="thaumatin-like protein; 9376-10898"
                     /protein_id="AAG52086.1"
                     /db_xref="GI:12324220"
                     /translation="MTNEKCEVALPLKLKSEKPSMASSSSLFLLPLLLLVSQASASTV
                     IFYNKCTYTVWPGIQPSSGQSLLAGGGFKLSPNRAYTLQLPPLWSGRFWGRHGCSFDR
                     SGRGRCATGDCGGSFLCNGAGGVPPATLAEITLGHDMDFYDVSLVDGYNLAMSIMPVK
                     GTGKCTYAGCVSDLNRMCPVGLQVRSRDGTQVVACKSACSAFNSPRYCCTGLFGNPQS
                     CKPTAYSKIFKVACPKAYSYAYDDPTSIATCSKANYVVTFCPHRGR"
     mRNA            <12692. .>13183
                     /gene="F25P22.4"
     gene            12692. .13183
                     /gene="F25P22.4"
                     /note="similar to calmodulin GB:P02596 [Renilla
                     reniformis]; contains Pfam profile: PF00036 EF hand (4
                     copies)"
     CDS             12692. .13183
                     /gene="F25P22.4"
                     /codon_start=1
                     /product="putative calmodulin; 12692-13183"
                     /protein_id="AAG52088.1"
                     /db_xref="GI:12324222"
                     /translation="MANTNLESTNKSTTPSTDMELKKVFDKFDANGDGKISVSELGNV
                     FKSMGTSYTEEELNRVLDEIDIDCDGFINQEEFATICRSSSSAVEIREAFDLYDQNKN
                     GLISSSEIHKVLNRLGMTCSVEDCVRMIGHVDTDGDGNVNFEEFQKMMSSPELVKGTV
                     ANS"
     mRNA            join(<14977. .15197,15451. .>15931)
                     /gene="F25P22.5"
     gene            14977. .15931
                     /gene="F25P22.5"
                     /note="contains Pfam profile: PF00071 ras family"
     CDS             join(14977. .15197,15451. .15931)
                     /gene="F25P22.5"
                     /codon_start=1
                     /product="putative ras-related GTP-binding protein;
                     14977-15931"
                     /protein_id="AAG52089.1"
                     /db_xref="GI:12324223"
                     /translation="MAEDTYEEECDYLFKAVLIGDSAVGKSNLLSRFSKDEFRFDSKP
                     TIGVEFAYRNVHVGDKIIKAQIWDTAGQERFRAITSSYYRGALGALLIYDITRRTTFD
                     NIKKWLFELRDFANPETVVVLVGNKSDLRQSREVEEDEGKTLAESEGLYFLETSALEN
                     VNVEEAFLVMIGRIHEVVTQRIASENKSNGAATPHINGNGNGTVLPVGKEIVNIHEVT
                     ATQPLLSSSSNCCFK"
     mRNA            complement(join(<16797. .16967,17064. .17297,17388. .17469,
                     17546. .17607,17706. .17773,17866. .17935,18448. .18772,
                     19179. .19223,19427. .>19725))
                     /gene="F25P22.7"
     gene            complement(16797. .19725)
                     /gene="F25P22.7"
     CDS             complement(join(16797. .16967,17064. .17297,17388. .17469,
                     17546. .17607,17706. .17773,17866. .17935,18448. .18772,
                     19179. .19223,19427. .19725))
                     /gene="F25P22.7"
                     /codon_start=1
                     /product="unknown protein; 19725-16797"
                     /protein_id="AAG52066.1"
                     /db_xref="GI:12324200"
                     /translation="MATLFTATVPSHHRFVSPSQHPKQSLLSQSLSVTFTENPQPTAV
                     VTLQEQQLTDWITSPVTRRFGIGAGFTWAGFLAFGVVSEQMKKSRLDVFQEEDNTRGL
                     EKQEEIILPNGIRYYDLQVGSGATPSSGYLVVFDVKGQVHGTEQVFVDTFGGKGKSLA
                     MVMDSRPYSKGLCQGIEHVLRSMKAGGKRKVIIPPSLGFGDRNVEFGQGLEIPPSATL
                     DYIIEVIYQFIFFVITALFKFDQVTDFAGSTNFVILAVLTLVLKASWHFRQIVLTLLV
                     VVWGLRLGIFLLMRILQWGEDRRFDEQRGNIVRLIIFWTLQAVWVWTVSLPLTLVNAS
                     DGGGSLKPADVIGWTMWVFGFLIEAAADQQKLSFKNSPENRGKWCDVGVWKYSRHPNY
                     FGEMLLWWGIFVAASPVLEGAEYLVIFGPLFLTLLLLFVSGIPLLEVLPCSLPLYLDN
                     F"
     mRNA            complement(join(<20191. .20329,20451. .20627,20721. .20770,
                     20870. .20938,21048. .21146,21239. .21317,21582. .21682,
                     21776. .21844,21957. .21997,22081. .23243,23382. .23485,
                     23568. .24008,24102. .>24662))
                     /gene="F25P22.8"
     gene            complement(20191. .24662)
                     /gene="F25P22.8"
                     /note="contains Pfam profile: PF00069 eukaryotic protein
                     kinase domain"
     CDS             complement(join(20191. .20329,20451. .20627,20721. .20770,
                     20870. .20938,21048. .21146,21239. .21317,21582. .21682,
                     21776. .21844,21957. .21997,22081. .23243,23382. .23485,
                     23568. .24008,24102. .24662))
                     /gene="F25P22.8"
                     /codon_start=1
                     /product="putative protein kinase; 24662-20191"
                     /protein_id="AAG52069.1"
                     /db_xref="GI:12324203"
                     /translation="MKVKEETLKNLGDGVVLRPVDHCSSIWSMKMNMKNFLKKLHISP
                     NQSDEAEGSISTTKSNHHKSIDVSSSSSPRSHHSNSPEIKPFSGLSNWLSSVGHRKIP
                     SPPNSFNAKNRAATVDDTVVVNGSEHVDLGSKDPAVEEENQIQLALELSAREDPEATQ
                     IEAIKQFSLGSCAPENSPAELIAYRYWNYNCLGYDDKILDGFYDLYGVLNASSAERIP

  Query Match             64.0%;  Score 16;  DB 12;  Length 105937;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   55362 AGTCACAGAATTATTT 55347


RESULT  33
AB020878
LOCUS       AB020878   110525 bp    DNA             PRI       21-MAY-1999
DEFINITION  Homo sapiens genomic DNA of 9q32 anti-oncogene of flat epitherium
            cancer , segment 10/10.
ACCESSION   AB020878
VERSION     AB020878.1  GI:4003398
KEYWORDS    .
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nakamura,Y., Isomura,M. and Daigo,Y.
  TITLE     Homo sapiens 1,010,525bp genomic DNA of 9q32 anti-oncogene of flat
            epitherium cancer region
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 110525)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1998) to the DDBJ/EMBL/GenBank databases. Mika
            Hirakawa, Japan Science and Technology Corporation (JST), Advanced
            Database Department; 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
            Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Japanese Foundation for Cancer
            Research as a JST seauencing team
            Principal Investigator: Yusuke Nakamura Ph.D.
            Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
            yusuke@hgc.ims.u-tokyo.ac.jp
            The sequence is submitted by:Human Genome Sequencing in ALIS
            project of JST
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further infomation about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site
            (http://www-alis.tokyo.jst.go.jp/HGS/top.html)
            or send email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES              Location/Qualifiers
     source           1. .110525
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="9"
                      /map="9q32"
     STS              80338. .80503
                      /standard_name="D9S1126"
                      /note="CHLC.GGAT11B01.P33342;The location is between each
```

```
                flanking site of PCR primers."
                /db_xref="GDB:686181"
     STS        86433. .86567
                /note="stSG8871;The location is between each flanking site
                of PCR primers."
                /db_xref="GDB:4569679"
     STS        104545. .104691
                /note="SHGC-18832;The location is between each flanking
                site of PCR primers."
                /db_xref="GDB:6458849"
BASE COUNT    32237 a  22817 c  23211 g  32215 t     45 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 85;  Length 110525;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   39879 AAGTCACAGAATTATT 39894


RESULT  34
AC011248
LOCUS       AC011248   111312 bp    DNA             PRI       28-APR-2000
DEFINITION  Homo sapiens 12 BAC RP11-14P21 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC011248
VERSION     AC011248.8  GI:7658299
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 111312)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
            Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
            Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W.,
            Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
            Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A.,
            Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
            Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
            Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
            Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
            Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
            Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
            Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
            Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 111312)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 111312)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28,. 2000 this sequence version replaced gi:6996901.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

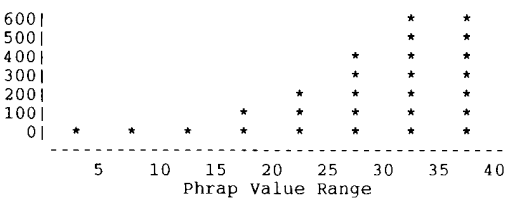
            QUALSTAT-REPORT----------


            ---------- Summary Statistics ------------------------------
            Contig length:                                  150286
            Phrap values in estimate:                       149476
            Average error rate (BCM-Phrap estimate):        0.000180006
            Fraction of Phrap values less than 40 :         0.0291418
            Number of consensus changing edits:             15
            Number of N's in consensus :                    1


            ---------- Consensus changing edits ------------------------
            Position      Original+Context          Edited+Context
            22441         aaagagaaga(n)gagggctcca   aaagagaaga(a)gagggctcca
            79231         agtaaggcca(n)annttttantn  agtaaggcca(t)atttttaatg
            79233         taaggccana(n)ntttantncc   taaggccata(t)ttttaatgcc
            79234         aaggccanan(n)tttantncct   aaggccatat(t)tttaatgcct
            79239         canannttta(n)tncctngtga   catattttta(a)tgcctcgtga
            79241         nannttttant(n)cctngtgaaa  tatttttaat(g)cctcgtgaaa
            79245         tttantncct(n)gtgaaaataa   tttaatgcct(c)gtgaaaataa
            79279         agagctgtga(n)gctagaaatg   agagctgtga(a)gctagaaatg
            79440         agcctggcca(n)catcgtgaaa   agcctggcca(a)catcgtgaaa
            89920         tctcaaaaaa(n)caaaaagaag   tctcaaaaaa(a)caaaaagaag
            94117         gtttttccag(n)tcctggcaca   gtttttccag(t)tcctggcaca
            95790         gcgagactcc(n)tctcaaaaaa   gcgagactcc(a)tctcaaaaaa
            111017        ttaaggttcc(n)tcagctgaac   ttaaggttcc(a)tcagctgaac
            115887        atgccatata(n)caacaaacaa   atgccatata(a)caacaaacaa
            129846        ggaggctgaa(n)cgagtggatc   ggaggctgaa(g)cgagtggatc


            --------- Distribution of Quality < 40 Bases ---------------------

                  1000|                                 *   *
                   900|                                 *   *
                   800|                                 *   *
              #    700|                                 *   *
```




```
          Version: 1.01  qxfo.
FEATURES             Location/Qualifiers
     source          1. .111312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-14P21"
     repeat_region   1. .133
                     /rpt_family="AluSg/x"
     repeat_region   167. .193
                     /rpt_family="(GAGAA)n"
     repeat_region   553. .795
                     /rpt_family="MIR"
     repeat_region   complement(824. .901)
                     /rpt_family="MIR"
     repeat_region   946. .985
                     /rpt_family="(CA)n"
     repeat_region   complement(1467. .1648)
                     /rpt_family="MLT1G"
     repeat_region   1876. .2176
                     /rpt_family="AluSq"
     repeat_region   complement(2180. .2431)
                     /rpt_family="AluSx"
     repeat_region   complement(2432. .2788)
                     /rpt_family="THE1B"
     repeat_region   complement(2789. .2835)
                     /rpt_family="AluSx"
     repeat_region   complement(2864. .3025)
                     /rpt_family="MLT1F1"
     repeat_region   3101. .3327
                     /rpt_family="MIR"
     repeat_region   complement(3383. .3455)
                     /rpt_family="L2"
     STS             3974. .4064
                     /standard_name="G07716"
                     /db_xref="dbSTS:13562"
     repeat_region   5179. .5480
                     /rpt_family="AluSp"
     repeat_region   complement(6278. .6347)
                     /rpt_family="MER4A"
     repeat_region   6348. .6648
                     /rpt_family="AluSq"
     repeat_region   complement(6649. .7184)
                     /rpt_family="MER4A"
     repeat_region   8486. .8703
                     /rpt_family="L2"
     repeat_region   8729. .9122
                     /rpt_family="L2"
     repeat_region   9213. .9497
                     /rpt_family="L2"
     repeat_region   9498. .9798
                     /rpt_family="AluY"
     repeat_region   9799. .9887
                     /rpt_family="L2"
     repeat_region   complement(10707. .12039)
                     /rpt_family="L1MA4"
     repeat_region   12047. .12155
                     /rpt_family="L2"
     repeat_region   complement(12165. .12729)
                     /rpt_family="L2"
     repeat_region   12702. .12847
                     /rpt_family="L2"
     STS             13332. .13486
                     /standard_name="Cda1fh02"
                     /db_xref="GDB:446262"
     repeat_region   13505. .13528

  Query Match             64.0%;  Score 16;  DB 87;  Length 111312;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   86830 AAGTCACAGAATTATT 86845


RESULT  35
AL590387
LOCUS       AL590387   118636 bp    DNA             HTG       08-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-356B20, *** SEQUENCING IN
            PROGRESS ***, 7 unordered pieces.
ACCESSION   AL590387
VERSION     AL590387.3  GI:13568351
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118636)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2001 this sequence version replaced gi:13561388.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ356B20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 116161 bases at least Q40
            Consensus quality: 117159 bases at least Q30
            Consensus quality: 117692 bases at least Q20
            Insert size: 118036; sum-of-contigs
            Insert size: 119811; 10.6% error; agarose-fp
            Quality coverage: 5.47x in Q20 bases; sum-of-contigs Quality
            coverage: 5.61x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    24283: contig of 24283 bp in length
            *    24284 24383: gap of      100 bp
            *    24384    26993: contig of 2610 bp in length
            *    26994 27093: gap of      100 bp
            *    27094    56307: contig of 29214 bp in length
            *    56308 56407: gap of      100 bp
            *    56408    59297: contig of 2890 bp in length
            *    59298 59397: gap of      100 bp
            *    59398    80206: contig of 20809 bp in length
            *    80207 80306: gap of      100 bp
            *    80307    97970: contig of 17664 bp in length
```

```
            *    97971 98070: gap of      100 bp
            *    98071   118636: contig of 20566 bp in length.
FEATURES             Location/Qualifiers
     source          1. .118636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-356B20"
                     /clone_lib="RPCI-3"
     misc_feature    1. .24283
                     /note="assembly_fragment:01399
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    24384. .26993
                     /note="assembly_fragment:00194
                     fragment_chain:1"
     misc_feature    27094. .56307
                     /note="assembly_fragment:00731
                     fragment_chain:1"
     misc_feature    56408. .59297
                     /note="assembly_fragment:00116
                     fragment_chain:1"
     misc_feature    59398. .80206
                     /note="assembly_fragment:00023
                     fragment_chain:1"
     misc_feature    80307. .97970
                     /note="assembly_fragment:00217
                     fragment_chain:1"
     misc_feature    98071. .118636
                     /note="assembly_fragment:00177
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    36924 a  21288 c  22673 g  37150 t    601 others
ORIGIN


  Query Match            64.0%;  Score 16;  DB 82;  Length 118636;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||||
Db   81402 AGTCACAGAATTATTT 81417


RESULT  36
AL159168
LOCUS       AL159168   129010 bp    DNA             PRI       30-SEP-2000
DEFINITION  Human DNA sequence from clone RP11-401H23 on chromosome 9, complete
            sequence.
ACCESSION   AL159168
VERSION     AL159168.15  GI:10443391
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129010)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:10178449.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-401H23 is from the library RPCI-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-401H23 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-401H23 is at 1 in this sequence.
            The true left end of clone RP11-16L21 is at 128911 in this
            sequence. The true right end of clone RP11-386D8 is at 19333 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .129010
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-401H23"
                     /clone_lib="RPCI-11.2"
BASE COUNT    39045 a  27222 c  26654 g  36089 t
ORIGIN


  Query Match            64.0%;  Score 16;  DB 90;  Length 129010;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||||
Db  102456 AGTCACAGAATTATTT 102471


RESULT  37
AC019289/c
LOCUS       AC019289   139218 bp    DNA             HTG       08-APR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-203I19 map 4, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC019289
VERSION     AC019289.3  GI:7528059
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-203I19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
```

```
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 8, 2000 this sequence version replaced gi:6778536.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4306
                Center clone name: 203_I_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 132555 bases at least Q40
                Consensus quality: 135869 bases at least Q30
                Consensus quality: 137053 bases at least Q20
                Insert size: 140000; agarose-fp
                Insert size: 138118; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      879: contig of 879 bp in length
            *      880 979: gap of      100 bp
            *      980     2319: contig of 1340 bp in length
            *     2320 2419: gap of      100 bp
            *     2420     6972: contig of 4553 bp in length
            *     6973 7072: gap of      100 bp
            *     7073    11233: contig of 4161 bp in length
            *    11234 11333: gap of      100 bp
            *    11334    18294: contig of 6961 bp in length
            *    18295 18394: gap of      100 bp
            *    18395    24919: contig of 6525 bp in length
            *    24920 25019: gap of      100 bp
            *    25020    32971: contig of 7952 bp in length
            *    32972 33071: gap of      100 bp
            *    33072    44845: contig of 11774 bp in length
            *    44846 44945: gap of      100 bp
            *    44946    60683: contig of 15738 bp in length
            *    60684 60783: gap of      100 bp
            *    60784    77274: contig of 16491 bp in length
            *    77275 77374: gap of      100 bp
            *    77375   102352: contig of 24978 bp in length
            *   102353 102452: gap of      100 bp
            *   102453   139218: contig of 36766 bp in length.
FEATURES             Location/Qualifiers
     source          1. .139218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-203I19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .879
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    980. .2319
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    2420. .6972
                     /note="assembly_fragment"
     misc_feature    7073. .11233
                     /note="assembly_fragment"
     misc_feature    11334. .18294
                     /note="assembly_fragment"
     misc_feature    18395. .24919
                     /note="assembly_fragment"
     misc_feature    25020. .32971
                     /note="assembly_fragment"
     misc_feature    33072. .44845
                     /note="assembly_fragment"
     misc_feature    44946. .60683
                     /note="assembly_fragment"
     misc_feature    60784. .77274
                     /note="assembly_fragment"
     misc_feature    77375. .102352
                     /note="assembly_fragment"
     misc_feature    102453. .139218
                     /note="assembly_fragment"
BASE COUNT    43319 a  25775 c  25817 g  43205 t   1102 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 65;  Length 139218;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db 103832 AGTCACAGAATTATTT 103817


RESULT  38
AC090105/c
LOCUS       AC090105   152452 bp    DNA             HTG       11-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-629E1 map 8, *** SEQUENCING IN
            PROGRESS ***, 51 unordered pieces.
ACCESSION   AC090105
VERSION     AC090105.2  GI:13273432
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152452)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-629E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152452)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
```

```
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 11, 2001 this sequence version replaced gi:12830262.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12772
                Center clone name: 629_E_1
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 51 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      654: contig of 654 bp in length
            *      655 754: gap of      100 bp
            *      755     1427: contig of 673 bp in length
            *     1428 1527: gap of      100 bp
            *     1528     2227: contig of 700 bp in length
            *     2228 2327: gap of      100 bp
            *     2328     3056: contig of 729 bp in length
            *     3057 3156: gap of      100 bp
            *     3157     3978: contig of 822 bp in length
            *     3979 4078: gap of      100 bp
            *     4079     4745: contig of 667 bp in length
            *     4746 4845: gap of      100 bp
            *     4846     6069: contig of 1224 bp in length
            *     6070 6169: gap of      100 bp
            *     6170     7193: contig of 1024 bp in length
            *     7194 7293: gap of      100 bp
            *     7294     8578: contig of 1285 bp in length
            *     8579 8678: gap of      100 bp
            *     8679     9808: contig of 1130 bp in length
            *     9809 9908: gap of      100 bp
            *     9909    11353: contig of 1445 bp in length
            *    11354 11453: gap of      100 bp
            *    11454    12575: contig of 1122 bp in length
            *    12576 12675: gap of      100 bp
            *    12676    14311: contig of 1636 bp in length
            *    14312 14411: gap of      100 bp
            *    14412    15641: contig of 1230 bp in length
            *    15642 15741: gap of      100 bp
            *    15742    16712: contig of 971 bp in length
            *    16713 16812: gap of      100 bp
            *    16813    18001: contig of 1189 bp in length
            *    18002 18101: gap of      100 bp
            *    18102    19709: contig of 1608 bp in length
            *    19710 19809: gap of      100 bp
            *    19810    21129: contig of 1320 bp in length
            *    21130 21229: gap of      100 bp
            *    21230    37074: contig of 15845 bp in length
            *    37075 37174: gap of      100 bp
            *    37175    39181: contig of 2007 bp in length
            *    39182 39281: gap of      100 bp
            *    39282    40867: contig of 1586 bp in length
            *    40868 40967: gap of      100 bp
            *    40968    42399: contig of 1432 bp in length
            *    42400 42499: gap of      100 bp
            *    42500    44459: contig of 1960 bp in length
            *    44460 44559: gap of      100 bp
            *    44560    47276: contig of 2717 bp in length
            *    47277 47376: gap of      100 bp
            *    47377    49419: contig of 2043 bp in length
            *    49420 49519: gap of      100 bp
            *    49520    51277: contig of 1758 bp in length
            *    51278 51377: gap of      100 bp
            *    51378    53690: contig of 2313 bp in length
            *    53691 53790: gap of      100 bp
            *    53791    56432: contig of 2642 bp in length
            *    56433 56532: gap of      100 bp
            *    56533    59175: contig of 2643 bp in length
            *    59176 59275: gap of      100 bp
            *    59276    61365: contig of 2090 bp in length
            *    61366 61465: gap of      100 bp
            *    61466    63403: contig of 1938 bp in length
            *    63404 63503: gap of      100 bp
            *    63504    66625: contig of 3122 bp in length
            *    66626 66725: gap of      100 bp
            *    66726    69388: contig of 2663 bp in length
            *    69389 69488: gap of      100 bp
            *    69489    71640: contig of 2152 bp in length
            *    71641 71740: gap of      100 bp
            *    71741    74645: contig of 2905 bp in length
            *    74646 74745: gap of      100 bp
            *    74746    77973: contig of 3228 bp in length
            *    77974 78073: gap of      100 bp
            *    78074    81010: contig of 2937 bp in length
            *    81011 81110: gap of      100 bp
            *    81111    83834: contig of 2724 bp in length
            *    83835 83934: gap of      100 bp
            *    83935    87888: contig of 3954 bp in length
            *    87889 87988: gap of      100 bp
            *    87989    91443: contig of 3455 bp in length
            *    91444 91543: gap of      100 bp
            *    91544    94541: contig of 2998 bp in length
            *    94542 94641: gap of      100 bp
            *    94642    99673: contig of 5032 bp in length
            *    99674 99773: gap of      100 bp
            *    99774   104302: contig of 4529 bp in length
            *   104303 104402: gap of      100 bp
            *   104403   109106: contig of 4704 bp in length
            *   109107 109206: gap of      100 bp
            *   109207   114261: contig of 5055 bp in length
            *   114262 114361: gap of      100 bp
            *   114362   120999: contig of 6638 bp in length
            *   121000 121099: gap of      100 bp
            *   121100   128145: contig of 7046 bp in length
            *   128146 128245: gap of      100 bp
            *   128246   133694: contig of 5449 bp in length
            *   133695 133794: gap of      100 bp
            *   133795   142736: contig of 8942 bp in length
            *   142737 142836: gap of      100 bp
            *   142837   151823: contig of 8987 bp in length
            *   151824 151923: gap of      100 bp
            *   151924   152452: contig of 529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .152452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-629E1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .654
                     /note="assembly_fragment"
     misc_feature    755. .1427
                     /note="assembly_fragment"
     misc_feature    1528. .2227
                     /note="assembly_fragment"
     misc_feature    2328. .3056
                     /note="assembly_fragment"
```

```
     misc_feature    3157. .3978
                     /note="assembly_fragment"
     misc_feature    4079. .4745
                     /note="assembly_fragment"
     misc_feature    4846. .6069
                     /note="assembly_fragment"
     misc_feature    6170. .7193
                     /note="assembly_fragment"
     misc_feature    7294. .8578
                     /note="assembly_fragment"
     misc_feature    8679. .9808
                     /note="assembly_fragment"
     misc_feature    9909. .11353
                     /note="assembly_fragment"
     misc_feature    11454. .12575
                     /note="assembly_fragment"
     misc_feature    12676. .14311
                     /note="assembly_fragment"
     misc_feature    14412. .15641
                     /note="assembly_fragment"
     misc_feature    15742. .16712
                     /note="assembly_fragment"
     misc_feature    16813. .18001

  Query Match             64.0%;  Score 16;  DB 77;  Length 152452;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db    4588 AGTCACAGAATTATTT 4573


RESULT  39
AC067998/c
LOCUS       AC067998   153221 bp    DNA             HTG       14-JUN-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-627D21 map 3, WORKING DRAFT
            SEQUENCE, 29 unordered pieces.
ACCESSION   AC067998
VERSION     AC067998.2  GI:8516111
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153221)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-627D21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153221)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 14, 2000 this sequence version replaced gi:7656755.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8575
                Center clone name: 627_D_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 137105 bases at least Q40
                Consensus quality: 145055 bases at least Q30
                Consensus quality: 148395 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 150421; sum-of-contigs
                Quality coverage: 3.9 in Q20 bases; agarose-fp
                Quality coverage: 4.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1623: contig of 1623 bp in length
            *     1624 1723: gap of      100 bp
            *     1724     3380: contig of 1657 bp in length
            *     3381 3480: gap of      100 bp
            *     3481     5348: contig of 1868 bp in length
            *     5349 5448: gap of      100 bp
            *     5449     7335: contig of 1887 bp in length
            *     7336 7435: gap of      100 bp
            *     7436     9535: contig of 2100 bp in length
            *     9536 9635: gap of      100 bp
            *     9636    12801: contig of 3166 bp in length
            *    12802 12901: gap of      100 bp
            *    12902    15358: contig of 2457 bp in length
            *    15359 15458: gap of      100 bp
            *    15459    17864: contig of 2406 bp in length
            *    17865 17964: gap of      100 bp
            *    17965    19498: contig of 1534 bp in length
            *    19499 19598: gap of      100 bp
            *    19599    21802: contig of 2204 bp in length
            *    21803 21902: gap of      100 bp
            *    21903    24102: contig of 2200 bp in length
            *    24103 24202: gap of      100 bp
            *    24203    26673: contig of 2471 bp in length
            *    26674 26773: gap of      100 bp
            *    26774    30361: contig of 3588 bp in length
            *    30362 30461: gap of      100 bp
            *    30462    34817: contig of 4356 bp in length
            *    34818 34917: gap of      100 bp
            *    34918    38602: contig of 3685 bp in length
            *    38603 38702: gap of      100 bp
            *    38703    43613: contig of 4911 bp in length
            *    43614 43713: gap of      100 bp
            *    43714    48515: contig of 4802 bp in length
            *    48516 48615: gap of      100 bp
            *    48616    53390: contig of 4775 bp in length
            *    53391 53490: gap of      100 bp
            *    53491    58254: contig of 4764 bp in length
            *    58255 58354: gap of      100 bp
            *    58355    65768: contig of 7414 bp in length
            *    65769 65868: gap of      100 bp
```

```
            *    65869    70737: contig of 4869 bp in length
            *    70738 70837: gap of      100 bp
            *    70838    77128: contig of 6291 bp in length
            *    77129 77228: gap of      100 bp
            *    77229    84888: contig of 7660 bp in length
            *    84889 84988: gap of      100 bp
            *    84989    94687: contig of 9699 bp in length
            *    94688 94787: gap of      100 bp
            *    94788   105941: contig of 11154 bp in length
            *   105942 106041: gap of      100 bp
            *   106042   113657: contig of 7616 bp in length
            *   113658 113757: gap of      100 bp
            *   113758   122759: contig of 9002 bp in length
            *   122760 122859: gap of      100 bp
            *   122860   136138: contig of 13279 bp in length
            *   136139 136238: gap of      100 bp
            *   136239   153221: contig of 16983 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-627D21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1623
                     /note="assembly_fragment"
     misc_feature    1724. .3380
                     /note="assembly_fragment"
     misc_feature    3481. .5348
                     /note="assembly_fragment"
     misc_feature    5449. .7335
                     /note="assembly_fragment"
     misc_feature    7436. .9535
                     /note="assembly_fragment"
     misc_feature    9636. .12801
                     /note="assembly_fragment"
     misc_feature    12902. .15358
                     /note="assembly_fragment"
     misc_feature    15459. .17864
                     /note="assembly_fragment"
     misc_feature    17965. .19498
                     /note="assembly_fragment"
     misc_feature    19599. .21802
                     /note="assembly_fragment"
     misc_feature    21903. .24102
                     /note="assembly_fragment"
     misc_feature    24203. .26673
                     /note="assembly_fragment"
     misc_feature    26774. .30361
                     /note="assembly_fragment"
     misc_feature    30462. .34817
                     /note="assembly_fragment"
     misc_feature    34918. .38602
                     /note="assembly_fragment"
     misc_feature    38703. .43613
                     /note="assembly_fragment"
     misc_feature    43714. .48515
                     /note="assembly_fragment"
     misc_feature    48616. .53390
                     /note="assembly_fragment"
     misc_feature    53491. .58254
                     /note="assembly_fragment"
     misc_feature    58355. .65768
                     /note="assembly_fragment"
     misc_feature    65869. .70737
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    70838. .77128
                     /note="assembly_fragment"
     misc_feature    77229. .84888
                     /note="assembly_fragment"
     misc_feature    84989. .94687
                     /note="assembly_fragment"
     misc_feature    94788. .105941
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    106042. .113657
                     /note="assembly_fragment"
     misc_feature    113758. .122759
                     /note="assembly_fragment"
     misc_feature    122860. .136138
                     /note="assembly_fragment"
     misc_feature    136239. .153221
                     /note="assembly_fragment"
BASE COUNT    42805 a  32811 c  32675 g  42124 t   2806 others
ORIGIN

  Query Match             64.0%;  Score 16;  DB 73;  Length 153221;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   14609 AGTCACAGAATTATTT 14594


RESULT  40
AC019258/c
LOCUS       AC019258   155211 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-2M23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC019258
VERSION     AC019258.2  GI:7209488
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155211)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-2M23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155211)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 8, 2000 this sequence version replaced gi:6649456.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2672
```

Center clone name: 2_M_23

* NOTE: This record contains 177 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
*        1      756: contig of 756 bp in length
*      757 856: gap of      100 bp
*      857     1619: contig of 763 bp in length
*     1620 1719: gap of      100 bp
*     1720     2498: contig of 779 bp in length
*     2499 2598: gap of      100 bp
*     2599     3355: contig of 757 bp in length
*     3356 3455: gap of      100 bp
*     3456     4201: contig of 746 bp in length
*     4202 4301: gap of      100 bp
*     4302     5053: contig of 752 bp in length
*     5054 5153: gap of      100 bp
*     5154     5910: contig of 757 bp in length
*     5911 6010: gap of      100 bp
*     6011     6779: contig of 769 bp in length
*     6780 6879: gap of      100 bp
*     6880     7629: contig of 750 bp in length
*     7630 7729: gap of      100 bp
*     7730     8485: contig of 756 bp in length
*     8486 8585: gap of      100 bp
*     8586     9348: contig of 763 bp in length
*     9349 9448: gap of      100 bp
*     9449    10225: contig of 777 bp in length
*    10226 10325: gap of      100 bp
*    10326    11083: contig of 758 bp in length
*    11084 11183: gap of      100 bp
*    11184    11964: contig of 781 bp in length
*    11965 12064: gap of      100 bp
*    12065    12830: contig of 766 bp in length
*    12831 12930: gap of      100 bp
*    12931    13692: contig of 762 bp in length
*    13693 13792: gap of      100 bp
*    13793    14557: contig of 765 bp in length
*    14558 14657: gap of      100 bp
*    14658    15412: contig of 755 bp in length
*    15413 15512: gap of      100 bp
*    15513    16282: contig of 770 bp in length
*    16283 16382: gap of      100 bp
*    16383    17131: contig of 749 bp in length
*    17132 17231: gap of      100 bp
*    17232    17980: contig of 749 bp in length
*    17981 18080: gap of      100 bp
*    18081    18839: contig of 759 bp in length
*    18840 18939: gap of      100 bp
*    18940    19701: contig of 762 bp in length
*    19702 19801: gap of      100 bp
*    19802    20565: contig of 764 bp in length
*    20566 20665: gap of      100 bp
*    20666    21415: contig of 750 bp in length
*    21416 21515: gap of      100 bp
*    21516    22286: contig of 771 bp in length
*    22287 22386: gap of      100 bp
*    22387    23161: contig of 775 bp in length
*    23162 23261: gap of      100 bp
*    23262    24022: contig of 761 bp in length
*    24023 24122: gap of      100 bp
*    24123    24872: contig of 750 bp in length
*    24873 24972: gap of      100 bp
*    24973    25654: contig of 682 bp in length
*    25655 25754: gap of      100 bp
*    25755    26510: contig of 756 bp in length
*    26511 26610: gap of      100 bp
*    26611    27369: contig of 759 bp in length
*    27370 27469: gap of      100 bp
*    27470    28209: contig of 740 bp in length
*    28210 28309: gap of      100 bp
*    28310    29077: contig of 768 bp in length
*    29078 29177: gap of      100 bp
*    29178    29924: contig of 747 bp in length
*    29925 30024: gap of      100 bp
*    30025    30796: contig of 772 bp in length
*    30797 30896: gap of      100 bp
*    30897    31655: contig of 759 bp in length
*    31656 31755: gap of      100 bp
*    31756    32502: contig of 747 bp in length
*    32503 32602: gap of      100 bp
*    32603    33369: contig of 767 bp in length
*    33370 33469: gap of      100 bp
*    33470    34233: contig of 764 bp in length
*    34234 34333: gap of      100 bp
*    34334    35081: contig of 748 bp in length
*    35082 35181: gap of      100 bp
*    35182    35951: contig of 770 bp in length
*    35952 36051: gap of      100 bp
*    36052    36820: contig of 769 bp in length
*    36821 36920: gap of      100 bp
*    36921    37694: contig of 774 bp in length
*    37695 37794: gap of      100 bp
*    37795    38563: contig of 769 bp in length
*    38564 38663: gap of      100 bp
*    38664    39427: contig of 764 bp in length
*    39428 39527: gap of      100 bp
*    39528    40285: contig of 758 bp in length
*    40286 40385: gap of      100 bp
*    40386    41148: contig of 763 bp in length
*    41149 41248: gap of      100 bp
*    41249    42007: contig of 759 bp in length
*    42008 42107: gap of      100 bp
*    42108    42867: contig of 760 bp in length
*    42868 42967: gap of      100 bp
*    42968    43716: contig of 749 bp in length
*    43717 43816: gap of      100 bp
*    43817    44549: contig of 733 bp in length
*    44550 44649: gap of      100 bp
*    44650    45404: contig of 755 bp in length
*    45405 45504: gap of      100 bp
*    45505    46270: contig of 766 bp in length
*    46271 46370: gap of      100 bp
*    46371    47132: contig of 762 bp in length
*    47133 47232: gap of      100 bp
*    47233    48006: contig of 774 bp in length
*    48007 48106: gap of      100 bp
*    48107    48875: contig of 769 bp in length
*    48876 48975: gap of      100 bp
*    48976    49735: contig of 760 bp in length
*    49736 49835: gap of      100 bp
*    49836    50585: contig of 750 bp in length
*    50586 50685: gap of      100 bp
*    50686    51449: contig of 764 bp in length
*    51450 51549: gap of      100 bp
*    51550    52306: contig of 757 bp in length
*    52307 52406: gap of      100 bp
*    52407    53158: contig of 752 bp in length
*    53159 53258: gap of      100 bp
*    53259    54002: contig of 744 bp in length
*    54003 54102: gap of      100 bp
*    54103    54883: contig of 781 bp in length
*    54884 54983: gap of      100 bp
*    54984    55748: contig of 765 bp in length
*    55749 55848: gap of      100 bp
*    55849    56621: contig of 773 bp in length
*    56622 56721: gap of      100 bp
*    56722    57477: contig of 756 bp in length
```

```
            *    57478 57577: gap of      100 bp
            *    57578    58335: contig of 758 bp in length
            *    58336 58435: gap of      100 bp
            *    58436    59188: contig of 753 bp in length
            *    59189 59288: gap of      100 bp
            *    59289    60050: contig of 762 bp in length
            *    60051 60150: gap of      100 bp
            *    60151    60905: contig of 755 bp in length
            *    60906 61005: gap of      100 bp

  Query Match             64.0%;  Score 16;  DB 65;  Length 155211;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   58944 AGTCACAGAATTATTT 58929


RESULT  41
AC041018/c
LOCUS       AC041018   156184 bp    DNA             HTG       22-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-478O20 map 4, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC041018
VERSION     AC041018.2  GI:8016453
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-478O20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7534220.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9577
                Center clone name: 478_O_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147958 bases at least Q40
                Consensus quality: 151874 bases at least Q30
                Consensus quality: 153413 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 154684; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1351: contig of 1351 bp in length
            *     1352 1451: gap of      100 bp
            *     1452     4369: contig of 2918 bp in length
            *     4370 4469: gap of      100 bp
            *     4470     8873: contig of 4404 bp in length
            *     8874 8973: gap of      100 bp
            *     8974    13167: contig of 4194 bp in length
            *    13168 13267: gap of      100 bp
            *    13268    18989: contig of 5722 bp in length
            *    18990 19089: gap of      100 bp
            *    19090    24989: contig of 5900 bp in length
            *    24990 25089: gap of      100 bp
            *    25090    33460: contig of 8371 bp in length
            *    33461 33560: gap of      100 bp
            *    33561    42328: contig of 8768 bp in length
            *    42329 42428: gap of      100 bp
            *    42429    50674: contig of 8246 bp in length
            *    50675 50774: gap of      100 bp
            *    50775    59862: contig of 9088 bp in length
            *    59863 59962: gap of      100 bp
            *    59963    71699: contig of 11737 bp in length
            *    71700 71799: gap of      100 bp
            *    71800    82583: contig of 10784 bp in length
            *    82584 82683: gap of      100 bp
            *    82684    95072: contig of 12389 bp in length
            *    95073 95172: gap of      100 bp
            *    95173   108949: contig of 13777 bp in length
            *   108950 109049: gap of      100 bp
            *   109050   126668: contig of 17619 bp in length
            *   126669 126768: gap of      100 bp
            *   126769   156184: contig of 29416 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-478O20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1351
                     /note="assembly_fragment"
     misc_feature    1452. .4369
                     /note="assembly_fragment"
     misc_feature    4470. .8873
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8974. .13167
                     /note="assembly_fragment"
     misc_feature    13268. .18989
                     /note="assembly_fragment"
     misc_feature    19090. .24989
                     /note="assembly_fragment"
     misc_feature    25090. .33460
```

```
                     /note="assembly_fragment"
     misc_feature    33561. .42328
                     /note="assembly_fragment"
     misc_feature    42429. .50674
                     /note="assembly_fragment"
     misc_feature    50775. .59862
                     /note="assembly_fragment"
     misc_feature    59963. .71699
                     /note="assembly_fragment"
     misc_feature    71800. .82583
                     /note="assembly_fragment"
     misc_feature    82684. .95072
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    95173. .108949
                     /note="assembly_fragment"
     misc_feature    109050. .126668
                     /note="assembly_fragment"
     misc_feature    126769. .156184
                     /note="assembly_fragment"
BASE COUNT    51507 a  26941 c  26653 g  49578 t   1505 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 71;  Length 156184;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatta 22
          ||||||||||||||||
Db 123532 GCAAGTCACAGAATTA 123517


RESULT  42
AL391821
LOCUS       AL391821   158408 bp    DNA             PRI       01-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-212B22 on chromosome X, complete
            sequence.
ACCESSION   AL391821
VERSION     AL391821.7  GI:11121082
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158408)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10944214.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-212B22 is from the library RPCI-11.1 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-212B22 The true
            left end of clone RP11-487I9 is at 133439 in this sequence. The
            true right end of clone RP11-35G18 is at 153997 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .158408
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-212B22"
                     /clone_lib="RPCI-11.1"
     repeat_region   2. .226
                     /note="L1MB2 repeat: matches 5946. .6163 of consensus"
     repeat_region   295. .525
                     /note="MLT1H repeat: matches 229. .494 of consensus"
     repeat_region   648. .677
                     /note="15 copies 2 mer ca 90% conserved"
     repeat_region   2598. .2649
                     /note="26 copies 2 mer ac 90% conserved"
     repeat_region   2780. .2853
                     /note="37 copies 2 mer ct 74% conserved"
     misc_feature    complement(3446. .3841)
                     /note="match: GSS: Em:AQ136371"
     repeat_region   4491. .4655
                     /note="L1ME repeat: matches 468. .628 of consensus"
     repeat_region   4804. .5008
                     /note="MER58C repeat: matches 3. .88 of consensus"
     repeat_region   5044. .5181
                     /note="L1M4 repeat: matches 1500. .1639 of consensus"
     repeat_region   5271. .5414
                     /note="L1MEc repeat: matches 2242. .2385 of consensus"
     repeat_region   5420. .5718
                     /note="AluSq repeat: matches 1. .298 of consensus"
     repeat_region   5726. .5994
                     /note="AluJo repeat: matches 1. .271 of consensus"
     repeat_region   6011. .6787
                     /note="L1MEc repeat: matches 2356. .2818 of consensus"
     repeat_region   6801. .7262
                     /note="L1PA5 repeat: matches 5682. .6143 of consensus"
     repeat_region   7305. .7750
                     /note="L1MA9 repeat: matches 5860. .6308 of consensus"
     repeat_region   7799. .8106
                     /note="L1M4 repeat: matches 2841. .3151 of consensus"
     repeat_region   8246. .8761
                     /note="L1M4 repeat: matches 869. .1405 of consensus"
     repeat_region   8819. .9030
                     /note="L1M4 repeat: matches 1391. .1596 of consensus"
     repeat_region   9073. .10246
                     /note="L1PA3 repeat: matches 4979. .6146 of consensus"
     repeat_region   10260. .10753
                     /note="L1M4 repeat: matches 1575. .1608 of consensus"
     repeat_region   11391. .11509
                     /note="FLAM_C repeat: matches 1. .132 of consensus"
     repeat_region   11510. .11821
                     /note="AluSq repeat: matches 1. .312 of consensus"
     repeat_region   11834. .12407
                     /note="L1M4 repeat: matches 2370. .2908 of consensus"
     repeat_region   12408. .12684
                     /note="AluJo repeat: matches 1. .277 of consensus"
     repeat_region   12685. .13112
                     /note="L1M4 repeat: matches 2908. .3346 of consensus"
     repeat_region   13191. .13493
                     /note="L1M4 repeat: matches 3350. .3653 of consensus"
     repeat_region   13515. .13538
                     /note="12 copies 2 mer at 100% conserved"
     repeat_region   13539. .13894
                     /note="L1MA9 repeat: matches 4974. .5334 of consensus"
     repeat_region   13895. .14205
                     /note="AluSg repeat: matches 1. .310 of consensus"
```

```
repeat_region   14206. .15193
                /note="L1MA9 repeat: matches 5334. .6305 of consensus"
repeat_region   15269. .15357
                /note="L2 repeat: matches 2644. .2748 of consensus"
repeat_region   15388. .15841
                /note="L1P repeat: matches 4685. .5138 of consensus"
misc_feature    complement(15793. .16200)
                /note="match: GSS: Em:AQ034620"
repeat_region   16099. .16221
                /note="L2 repeat: matches 2572. .2710 of consensus"
repeat_region   16280. .16732
                /note="L2 repeat: matches 1992. .2403 of consensus"
repeat_region   17982. .18300
                /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region   18532. .18756
                /note="MIR repeat: matches 14. .262 of consensus"
repeat_region   20534. .20587
                /note="27 copies 2 mer tt 81% conserved"
repeat_region   21039. .21171
                /note="AluSg/x repeat: matches 79. .209 of consensus"
repeat_region   22602. .22902
                /note="AluSx repeat: matches 2. .302 of consensus"
repeat_region   23484. .23533
                /note="25 copies 2 mer gt 96% conserved"
repeat_region   23532. .23561
                /note="10 copies 3 mer gtg 90% conserved"
repeat_region   23877. .25040
                /note="L2 repeat: matches 1141. .2323 of consensus"
repeat_region   25257. .25415
                /note="L2 repeat: matches 2582. .2748 of consensus"
repeat_region   26553. .26594
                /note="21 copies 2 mer aa 76% conserved"
misc_feature    26798. .27063
                /note="match: GSS: Em:AQ343061"
repeat_region   27948. .28106
                /note="MER5B repeat: matches 10. .175 of consensus"
repeat_region   28430. .28618
                /note="MER53 repeat: matches 1. .189 of consensus"
repeat_region   28941. .29026
                /note="MIR repeat: matches 112. .198 of consensus"
repeat_region   30445. .30809
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   31228. .31341
                /note="L2 repeat: matches 2629. .2749 of consensus"
repeat_region   31996. .32079
                /note="L2 repeat: matches 2620. .2698 of consensus"
repeat_region   32105. .32398
                /note="AluY repeat: matches 1. .294 of consensus"
misc_feature    complement(32313. .32867)
                /note="match: GSS: Em:AQ357777"
misc_feature    complement(32386. .32876)
                /note="match: GSS: Em:AQ357649"
repeat_region   32611. .32654
                /note="22 copies 2 mer tt 79% conserved"
repeat_region   32804. .32877
                /note="MER5A repeat: matches 21. .114 of consensus"
repeat_region   33673. .33723
                /note="MIR repeat: matches 102. .152 of consensus"
misc_feature    complement(34512. .34778)
                /note="match: STS: Em:L24830"
repeat_region   35370. .35678
                /note="AluY repeat: matches 1. .309 of consensus"
misc_feature    complement(36574. .36908)
                /note="match: STS: Em:HSC93A6"
repeat_region   36808. .37545
                /note="L2 repeat: matches 2017. .2749 of consensus"
repeat_region   37906. .38033
                /note="MER5A repeat: matches 59. .189 of consensus"
repeat_region   38390. .38732
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region   38744. .38995
                /note="MIR repeat: matches 8. .261 of consensus"
misc_feature    complement(39579. .40134)
                /note="match: STS: Em:HSC18H06"
repeat_region   39990. .40124
                /note="L2 repeat: matches 2564. .2710 of consensus"
repeat_region   42729. .42884
                /note="MER5A repeat: matches 4. .184 of consensus"
repeat_region   43123. .43367
                /note="MIR repeat: matches 3. .261 of consensus"
repeat_region   44086. .44327
                /note="AluSg/x repeat: matches 70. .308 of consensus"
repeat_region   45117. .45305
                /note="MER33 repeat: matches 1. .194 of consensus"
repeat_region   45425. .45951
                /note="L2 repeat: matches 1767. .2358 of consensus"
repeat_region   46025. .46336
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   46474. .46532
                /note="MIR repeat: matches 65. .123 of consensus"

  Query Match             64.0%;  Score 16;  DB 90;  Length 158408;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
            ||||||||||||||||
Db 117994 AAGTCACAGAATTATT 118009


RESULT  43
AC022621
LOCUS       AC022621   167110 bp    DNA             HTG       14-MAR-2000
DEFINITION  Homo sapiens clone RP11-23M2, WORKING DRAFT SEQUENCE, 27 unordered
            pieces.
ACCESSION   AC022621
VERSION     AC022621.4  GI:7239605
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167110)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-23M2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167110)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 14, 2000 this sequence version replaced gi:6980282.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
```

```
-------------- Project Information
    Center project name: L4396
    Center clone name: 23_M_2
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 154068 bases at least Q40
    Consensus quality: 160659 bases at least Q30
    Consensus quality: 162957 bases at least Q20
    Insert size: 152000; agarose-fp
    Insert size: 164510; sum-of-contigs
    Quality coverage: 4.2 in Q20 bases; agarose-fp
    Quality coverage: 3.9 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     1024: contig of 1024 bp in length
*    1025  1124: gap of      100 bp
*    1125     2237: contig of 1113 bp in length
*    2238  2337: gap of      100 bp
*    2338     3378: contig of 1041 bp in length
*    3379  3478: gap of      100 bp
*    3479     5102: contig of 1624 bp in length
*    5103  5202: gap of      100 bp
*    5203     7984: contig of 2782 bp in length
*    7985  8084: gap of      100 bp
*    8085    10770: contig of 2686 bp in length
*   10771 10870: gap of      100 bp
*   10871    13246: contig of 2376 bp in length
*   13247 13346: gap of      100 bp
*   13347    15161: contig of 1815 bp in length
*   15162 15261: gap of      100 bp
*   15262    18865: contig of 3604 bp in length
*   18866 18965: gap of      100 bp
*   18966    22975: contig of 4010 bp in length
*   22976 23075: gap of      100 bp
*   23076    26777: contig of 3702 bp in length
*   26778 26877: gap of      100 bp
*   26878    31543: contig of 4666 bp in length
*   31544 31643: gap of      100 bp
*   31644    36119: contig of 4476 bp in length
*   36120 36219: gap of      100 bp
*   36220    40123: contig of 3904 bp in length
*   40124 40223: gap of      100 bp
*   40224    42742: contig of 2519 bp in length
*   42743 42842: gap of      100 bp
*   42843    49609: contig of 6767 bp in length
*   49610 49709: gap of      100 bp
*   49710    55228: contig of 5519 bp in length
*   55229 55328: gap of      100 bp
*   55329    60561: contig of 5233 bp in length
*   60562 60661: gap of      100 bp
*   60662    67385: contig of 6724 bp in length
*   67386 67485: gap of      100 bp
*   67486    73407: contig of 5922 bp in length
*   73408 73507: gap of      100 bp
*   73508    81201: contig of 7694 bp in length
*   81202 81301: gap of      100 bp
*   81302    89851: contig of 8550 bp in length
*   89852 89951: gap of      100 bp
*   89952    97520: contig of 7569 bp in length
*   97521 97620: gap of      100 bp
*   97621   110089: contig of 12469 bp in length
*  110090 110189: gap of      100 bp
*  110190   122757: contig of 12568 bp in length
*  122758 122857: gap of      100 bp
*  122858   143904: contig of 21047 bp in length
*  143905 144004: gap of      100 bp
*  144005   167110: contig of 23106 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167110
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-23M2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1024
                     /note="assembly_fragment"
     misc_feature    1125. .2237
                     /note="assembly_fragment"
     misc_feature    2338. .3378
                     /note="assembly_fragment"
     misc_feature    3479. .5102
                     /note="assembly_fragment"
     misc_feature    5203. .7984
                     /note="assembly_fragment"
     misc_feature    8085. .10770
                     /note="assembly_fragment"
     misc_feature    10871. .13246
                     /note="assembly_fragment"
     misc_feature    13347. .15161
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    15262. .18865
                     /note="assembly_fragment"
     misc_feature    18966. .22975
                     /note="assembly_fragment"
     misc_feature    23076. .26777
                     /note="assembly_fragment"
     misc_feature    26878. .31543
                     /note="assembly_fragment"
     misc_feature    31644. .36119
                     /note="assembly_fragment"
     misc_feature    36220. .40123
                     /note="assembly_fragment"
     misc_feature    40224. .42742
                     /note="assembly_fragment"
     misc_feature    42843. .49609
                     /note="assembly_fragment"
     misc_feature    49710. .55228
                     /note="assembly_fragment"
     misc_feature    55329. .60561
                     /note="assembly_fragment"
     misc_feature    60662. .67385
                     /note="assembly_fragment"
     misc_feature    67486. .73407
                     /note="assembly_fragment"
     misc_feature    73508. .81201
                     /note="assembly_fragment"
     misc_feature    81302. .89851
                     /note="assembly_fragment"
     misc_feature    89952. .97520
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    97621. .110089
                     /note="assembly_fragment"
     misc_feature    110190. .122757
                     /note="assembly_fragment"
     misc_feature    122858. .143904
                     /note="assembly_fragment"
     misc_feature    144005. .167110
                     /note="assembly_fragment"
BASE COUNT    48189 a  34563 c  33851 g  47895 t   2612 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 67;  Length 167110;
  Best Local Similarity   100.0%;  Pred. No. 12;
```

```
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||
Db   49204 AGTCACAGAATTATTT 49219


RESULT  44
AC024560
LOCUS       AC024560   167176 bp    DNA             HTG       16-APR-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-496H1, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   AC024560
VERSION     AC024560.15  GI:13624340
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167176)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167176)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 14, 2001 this sequence version replaced gi:13374578.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAJW
                Center clone name: RP11-496H1
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 6% of reads
                Chemistry: Dye-terminator Big Dye: 94% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162785 bases at least Q40
                Consensus quality: 165676 bases at least Q30
                Consensus quality: 167353 bases at least Q20
                Estimated insert size: 164567; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    69736: contig of 69736 bp in length
            *    69737    69836: gap of unknown length
            *    69837   101725: contig of 31889 bp in length
            *   101726   101825: gap of unknown length
            *   101826   122611: contig of 20786 bp in length
            *   122612   122711: gap of unknown length
            *   122712   139412: contig of 16701 bp in length
            *   139413   139512: gap of unknown length
            *   139513   154634: contig of 15122 bp in length
            *   154635   154734: gap of unknown length
            *   154735   164776: contig of 10042 bp in length
            *   164777   164876: gap of unknown length
            *   164877   166032: contig of 1156 bp in length
            *   166033   166132: gap of unknown length
            *   166133   167176: contig of 1044 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167176
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-496H1"
BASE COUNT    45856 a  40392 c  37326 g  42895 t    707 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 68;  Length 167176;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||
Db 167115 AGTCACAGAATTATTT 167130


RESULT  45
AC010816
LOCUS       AC010816   167803 bp    DNA             HTG       10-SEP-2000
DEFINITION  Homo sapiens clone RP11-3G13, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC010816
VERSION     AC010816.4  GI:10047737
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167803)
```

```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-3G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167803)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:7637262.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2730
                Center clone name: 3_G_13
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150949 bases at least Q40
                Consensus quality: 159528 bases at least Q30
                Consensus quality: 163392 bases at least Q20
                Insert size: 145000; agarose-fp
                Insert size: 166803; sum-of-contigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp
                Quality coverage: 3.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1593: contig of 1593 bp in length
            *     1594  1693: gap of      100 bp
            *     1694     2814: contig of 1121 bp in length
            *     2815  2914: gap of      100 bp
            *     2915    47564: contig of 44650 bp in length
            *    47565 47664: gap of      100 bp
            *    47665    59530: contig of 11866 bp in length
            *    59531 59630: gap of      100 bp
            *    59631    70197: contig of 10567 bp in length
            *    70198 70297: gap of      100 bp
            *    70298    81312: contig of 11015 bp in length
            *    81313 81412: gap of      100 bp
            *    81413    99699: contig of 18287 bp in length
            *    99700 99799: gap of      100 bp
            *    99800   116730: contig of 16931 bp in length
            *   116731 116830: gap of      100 bp
            *   116831   139677: contig of 22847 bp in length
            *   139678 139777: gap of      100 bp
            *   139778   165369: contig of 25592 bp in length
            *   165370 165469: gap of      100 bp
            *   165470   167803: contig of 2334 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167803
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-3G13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1593
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1694. .2814
                     /note="assembly_fragment"
     misc_feature    2915. .47564
                     /note="assembly_fragment"
     misc_feature    47665. .59530
                     /note="assembly_fragment"
     misc_feature    59631. .70197
                     /note="assembly_fragment"
     misc_feature    70298. .81312
                     /note="assembly_fragment"
     misc_feature    81413. .99699
                     /note="assembly_fragment"
     misc_feature    99800. .116730
                     /note="assembly_fragment"
     misc_feature    116831. .139677
                     /note="assembly_fragment"
     misc_feature    139778. .165369
                     /note="assembly_fragment"
     misc_feature    165470. .167803
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54453 a  30838 c  30293 g  51217 t   1002 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 61;  Length 167803;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   83723 AGTCACAGAATTATTT 83738


Search completed: October  9, 2001, 15:44:37
Job time: 14663 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:51:43 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 27371

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	20	AAX01303	E. coli biotin syn
	2	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	3	25	100.0	1121	7	AAN60496	Sequence encoding
	4	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
c	5	17	68.0	17	14	AAQ56307	BioB DNA primer.
	6	15	60.0	561	22	AAF94235	Primer specific fo
	7	15	60.0	1771	20	AAZ30602	Mouse integrin alp
c	8	15	60.0	3520	22	AAF93909	Human cDNA encodin
c	9	15	60.0	3520	22	AAF29361	Amyloid-beta prote
c	10	15	60.0	116624	19	AAV52850	Human eyal gene co
	11	14	56.0	329	21	AAA41925	Human secreted exp
	12	14	56.0	360	22	AAF65421	Novel human polynu
	13	14	56.0	385	14	AAQ39649	Expressed Sequence
	14	14	56.0	385	14	AAQ59061	Human brain Expres
c	15	14	56.0	902	21	AAZ56756	Human transmembran
c	16	14	56.0	1050	9	AAN80927	Lividans Exported
c	17	14	56.0	1677	17	AAT31932	Retinoid X recepto
c	18	14	56.0	1787	17	AAT31928	Retinoid X recepto
c	19	14	56.0	1890	20	AAZ52940	Human prostate tum
c	20	14	56.0	2233	19	AAV10120	Human retinoid rec
c	21	14	56.0	2585	22	AAF77541	Caenorhabditis bri
c	22	14	56.0	2845	21	AAC69462	Human secreted pro
c	23	14	56.0	3413	20	AAV72330	F. balustinum CP70
	24	14	56.0	5684	21	AAC59953	Human secreted pro
c	25	14	56.0	5773	21	AAC76014	Human ORFX ORF1569
	26	14	56.0	8041	17	AAT10105	Adherence conferri
	27	14	56.0	32351	21	AAF21307	Human low adenosin
	28	14	56.0	32351	21	AAA35185	Human adenosine re
	29	14	56.0	40298	21	AAF21311	Human low adenosin
	30	14	56.0	40298	21	AAA35189	Human adenosine re
	31	14	56.0	81001	22	AAF30035	Human apolipoprote
	32	14	56.0	1038602	20	AAZ01425	Complete genome se
	33	13	52.0	87	16	AAT22576	Human gene signatu
	34	13	52.0	105	16	AAT20380	Human gene signatu
c	35	13	52.0	167	16	AAT20233	Human gene signatu
c	36	13	52.0	187	21	AAC26213	Human secreted pro
	37	13	52.0	354	15	AAQ77082	Human genome fragm
c	38	13	52.0	363	21	AAA50159	Antibody 5H7 heavy
c	39	13	52.0	378	13	AAQ25667	Sequence of the an
	40	13	52.0	387	21	AAC79285	Human lung tumour-
c	41	13	52.0	408	21	AAA74604	Murine 5B3 antibod
	42	13	52.0	409	21	AAC03836	Human secreted pro
c	43	13	52.0	415	21	AAC56783	Eucalyptus grandis
c	44	13	52.0	441	21	AAC07649	Human secreted pro
	45	13	52.0	476	21	AAC07951	Human secreted pro

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 attgtcgcaagtcacagaattattt 48


RESULT   2
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     47 attgtcgcaagtcacagaattattt 71


RESULT   3
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     65 attgtcgcaagtcacagaattattt 89


RESULT   4
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
```

```
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 attgtcgcaagtcacagaattattt 164


RESULT   5
AAQ56307/c
ID   AAQ56307 standard; DNA; 17 BP.
XX
AC   AAQ56307;
XX
DT   13-APR-1994  (first entry)
XX
DE   BioB DNA primer.
XX
KW   BioA; BioB; promoter; biotin; operon; primer; ss.
XX
OS   Synthetic.
XX
PN   JP05219956-A.
```

XX
PD 31-AUG-1993.
XX
PF 14-SEP-1992; 92JP-0244792.
XX
PR 14-SEP-1992; 92JP-0244792.
XX
PA (SHIS) SHISEIDO CO LTD.
PA (TAKE) TAKEDA CHEM IND LTD.
XX
DR WPI; 1993-308323/39.
XX
PT DNA sequence of biotin operon - has base sequence of E. coli
PT mutated by base pair(s) compared to wild type
XX
PS Example 1; Fig 8; 11pp; Japanese.
XX
CC A novel DNA sequence comprises the E.coli biotin operon (BO) in which
CC the control region of BO or the region near the bioB initiation
CC codon is mutated by at least one base pair compared to its
CC wild type. Two primers (AAQ56306-Q56307) are described in Example 1.
CC A microorganism belonging to Escherichia genus, transformed by
CC a recombinant plasmid carrying such DNA can be used for the prodn.
CC of biotin-active substances.
XX
SQ Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gtcgcaagtcacagaat 20
||||||||||||||||
Db 17 GTCGCAAGTCACAGAAT 1

RESULT 6
AAF94235
ID AAF94235 standard; DNA; 561 BP.
XX
AC AAF94235;
XX
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS Synthetic.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 5; SEQ ID 669; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;

Query Match 60.0%; Score 15; DB 22; Length 561;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
|||||||||||||||
Db 339 caagtcacagaatta 353

RESULT 7
AAZ30602
ID AAZ30602 standard; DNA; 1771 BP.
XX
AC AAZ30602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW vascular endothelial cell; vascular endothelium; migration; inflammation;
KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW metastasis; circulatory system; autoimmune disease; Grave's disease;
KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998; 98US-0166203.
XX
PR 05-OCT-1998; 98US-0166203.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowsert LM, Condon TP;
XX
DR WPI; 1999-590416/50.
DR P-PSDB; AAY44154.
XX
PT Antisense inhibition of integrin alpha4 expression useful for treating
PT inflammatory diseases such as atherosclerosis, allergies, asthma and

```
PT   arthritis -
XX
PS   Example 12; Column 71-74; 40pp; English.
XX
CC   This sequence represents the coding region of the mouse integrin
CC   alpha4 gene. The invention relates to the generation of antisense
CC   oligonucleotides targeted to the integrin alpha4 gene which are used
CC   for inhibiting expression of the integrin alpha4 mRNA or protein.
CC   Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
CC   called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
CC   surfaces of leucocytes and vascular endothelial cells and mediates the
CC   adhesion of leucocytes to the vascular endothelium prior to migration
CC   into the surrounding tissues. This migration is an essential step in
CC   inflammation and hence VLA-4 (and consequently integrin alpha4) is a
CC   potential therapeutic target for treating inflammatory diseases and
CC   the damaging effects of excessive inflammation. These disorders include
CC   atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
CC   metastasis (VLA-4 is involved in migration of the tumor cells through
CC   the extracellular matrix into the circulatory system). VLA-4 is also
CC   involved in a number of autoimmune diseases such as Grave's disease,
CC   Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.
CC   VLA-4 may also be involved in promoting adhesion (i.e. retention) of
CC   hemopoietic stem cells in bone-marrow and in allograft rejection.
XX
SQ   Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   8
AAF93909/c
ID   AAF93909 standard; cDNA; 3520 BP.
XX
AC   AAF93909;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0256.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88482.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 331; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           |||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT   9
AAF29361/c
ID   AAF29361 standard; DNA; 3520 BP.
XX
AC   AAF29361;
XX
DT   20-APR-2001  (first entry)
XX
DE   Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
XX
KW   Human; amyloid-beta protein; agglutination regulatory factor;
KW   Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200104299-A1.
XX
PD   18-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-JP04515.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   18-OCT-1999;   99US-0159586.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Yamazaki M,  Satoh S;
PI   Arakawa H,  Morita M;
XX
DR   WPI; 2001-138347/14.
DR   P-PSDB; AAB49771.
XX
PT   Polynucleotide encoding Amyloid-beta protein agglutination-controlling
PT   factor, useful for inhibiting or promoting agglutination or
PT   sedimentation of amyloid-beta protein and in diagnosis and screening
PT   drugs for Alzheimer's disease -
```

XX
PS Claim 1; Page 58-63; 72pp; Japanese.
XX
CC This invention relates to polynucleotides AAF29357 - AAF29361 which
CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
CC agglutination of amyloid beta protein. The protein and polynucleotide
CC sequences are useful in the diagnosis of Alzheimer's disease. They are
CC also useful for screening drugs which are useful for treating Alzheimer's
CC disease.
XX
SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
||||||||||||||||
Db 3182 CAAGTCACAGAATTA 3168

RESULT 10
AAV52850/c
ID AAV52850 standard; DNA; 116624 BP.
XX
AC AAV52850;
XX
DT 06-NOV-1998 (first entry)
XX
DE Human eya1 gene contig 4405-9480.
XX
KW Human; eya1; branchio-oto-renal syndrome; BOR; carcinogenesis; cancer;
KW kidney tumourogenesis; tumour; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9832849-A2.
XX
PD 30-JUL-1998.
XX
PF 28-JAN-1998; 98WO-EP00433.
XX
PR 29-JAN-1997; 97US-0036579.
XX
PA (INSP) INST PASTEUR.
XX
PI Abdelhak S, Compain S, Petit C, Vasiliki K, Vincent C;
PI Weil D;
XX
DR WPI; 1998-427945/36.
XX
PT Nucleic acid corresponding to human genes implicated in
PT branchio-oto-renal syndrome - useful for, e.g. diagnosis and
PT treatment of syndrome and possibly some carcinogenic processes,
PT particularly in kidneys
XX
PS Claim 1; Page 44-105; 191pp; English.
XX
CC The present sequence represents a portion of the human eya-1 gene.
CC Alterations in the eya1 gene are associated with branchio-oto-renal
CC (BOR) syndrome, including the form (BO) without renal anomalies, and
CC possibly in some carcinogenic processes, particularly in the kidney.
CC The related genes eya-2 and eya-3 may also be involved in development
CC of tumours and cell differentiation. Antibodies specific for EYA
CC proteins are used to detect the proteins by immunoassay, while genetic
CC alterations linked to BOR are detected by amplifying DNA with primers
CC (see AAV52857 to AAV52893) then identifying any mutations or deletions.
CC EYA1 and EYA1-B proteins, corresponding nucleic acid, antisense sequences
CC and Ab can be used therapeutically to modulate expression of EYA1(B), or
CC their active derivatives, especially for treating renal disease
CC associated with eya1 abnormalities, also for ensuring correct
CC development of grafted organs, especially where these are embryonic.
CC Therapeutic proteins are administered to maintain 0.1-10 mu g /ml in
CC body fluids, locally or systemically. Nucleic acid is administered in
CC usual gene therapy vectors (optionally to cells ex vivo) at doses of
CC 0.1-100 mu g.
XX
SQ Sequence 116624 BP; 35549 A; 20868 C; 23043 G; 37152 T; 12 other;

Query Match 60.0%; Score 15; DB 19; Length 116624;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 gtcacagaattattt 25
|||||||||||||||
Db 19735 GTCACAGAATTATTT 19721

RESULT 11
AAA41925
ID AAA41925 standard; cDNA; 329 BP.
XX
AC AAA41925;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:665.
XX
KW Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021990-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
DR WPI; 2000-317937/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 308; 618pp; English.
XX
CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;

CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 329 BP; 97 A; 84 C; 67 G; 81 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 329;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattattt 25
||||||||||||||
Db 198 tcacagaattattt 211

RESULT 12
AAF65421
ID AAF65421 standard; cDNA; 360 BP.
XX
AC AAF65421;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 1177.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 712; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 360 BP; 95 A; 70 C; 69 G; 125 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 360;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 gtcacagaattatt 24
||||||||||||||
Db 71 gtcacagaattatt 84

RESULT 13
AAQ39649
ID AAQ39649 standard; DNA; 385 BP.
XX
AC AAQ39649;
XX
DT 20-MAY-1993 (first entry)
XX
DE Expressed Sequence Tag human gene marker EST00025.
XX
KW expressed sequence tag; human genome project; chromosome;
KW human gene sequencing; PCR mapping; somatic cell hybrids;
KW sublocalisation; gene tagging; tissue typing.
XX
OS Synthetic.
XX
PN WO9300353-A.
XX
PD 07-JAN-1993.
XX
PF 19-JUN-1992; 92WO-US05222.
XX
PR 20-JUN-1991; 91US-0716831.
PR 12-FEB-1992; 92US-0837195.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
PI Adams MD, Venter JC;
XX
DR WPI; 1993-036325/04.
XX
PT Particular expressed sequence tags from human CDNA - corresponds
PT to transcription prods. of genes, useful for tagging genes,
PT mapping chromosomes and tissue typing
XX
PS Claim 3; Page 86; 199pp; English.
XX
CC This sequence represents an EST (expressed sequence tag) ESTs are
CC markers for human genes actually transcribed in vivo. Unlike the random
CC genomic DNA sequence tagged sites (STSs), ESTs point directly to
CC expressed genes. The use of ESTs could facilitate the tagging of most
CC expressed human genes within a few years at a fraction of the cost of
CC complete genomic sequencing. Using PCR primers AAQ39419-Q39580 (sequences
CC designed from the ESTs) sublocalisation of an EST can be achieved with
CC panels of fragments from specific chromosomes or pools of large genomic
CC clones in an analogous manner. This sequence represents EST00025.
XX
SQ Sequence 385 BP; 101 A; 66 C; 78 G; 139 T; 1 other;

```
  Query Match            56.0%;  Score 14;  DB 14;  Length 385;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           |||||||||||||||
Db     116 tcacagaattattt 129


RESULT  14
AAQ59061
ID   AAQ59061 standard; cDNA; 385 BP.
XX
AC   AAQ59061;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST00025.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 1; Page 115; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST00025 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM and hybridises to chromosome 20.
CC   See also AAQ59041-Q61440.
XX
SQ   Sequence 385 BP; 99 A; 66 C; 78 G; 140 T; 2 other;


  Query Match            56.0%;  Score 14;  DB 14;  Length 385;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           |||||||||||||||
Db     116 tcacagaattattt 129


RESULT  15
AAZ56756/c
ID   AAZ56756 standard; cDNA; 902 BP.
XX
AC   AAZ56756;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human transmembrane protein HTMPN-59 encoding cDNA.
XX
KW   Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW   antiproliferative; neuroprotective; immune disorder;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   gastrointestinal disorder; developmental disorder;
KW   cell proliferative disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO9961471-A2.
XX
PD   02-DEC-1999.
XX
PF   28-MAY-1999;   99WO-US11904.
XX
PR   29-MAY-1998;   98US-0087260.
PR   02-JUL-1998;   98US-0091674.
PR   02-OCT-1998;   98US-0102954.
PR   24-NOV-1998;   98US-0109869.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Hillman JL,  Yue H,  Guegler KJ,  Corley NC;
PI   Bandman O,  Patterson C,  Gorgone GA,  Kaser MR,  Baughn MR;
PI   Au-Young J;
XX
DR   WPI; 2000-072605/06.
DR   P-PSDB; AAY57935.
XX
PT   Proteins, polynucleotides, vectors, host cells and antibodies used to
PT   diagnose, treat or prevent immune, reproductive, smooth muscle,
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders -
XX
PS   Claim 9; Page 215; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neurological,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;


  Query Match            56.0%;  Score 14;  DB 21;  Length 902;
  Best Local Similarity  100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           |||||||||||||||
Db     425 AGTCACAGAATTAT 412


RESULT  16
AAN80927/c
ID   AAN80927 standard; DNA; 1050 BP.
XX
AC   AAN80927;
XX
DT   15-OCT-1990  (first entry)
XX
DE   Lividans Exported Protein gene.
XX
KW   Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
```

```
KW   protease inhibitor; probe; ss.
XX
OS   Streptomyces lividans strain 1326.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     394..498
FT                   /*tag=  a
FT   mat_peptide     499.. 825
FT                   /*tag=  b
FT                   /product=LEP-10
XX
PN   WO8801278-A.
XX
PD   25-FEB-1988.
XX
PF   17-AUG-1987;   87WO-US02009.
XX
PR   18-AUG-1986;   86US-0897245.
XX
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PA   (BERK/) TR BERKA.
XX
PI   Berka TR,  Fornwald JA,  Gorniak JG,  Rosenberg M;
XX
DR   WPI; 1988-063996/09.
DR   P-PSDB; AAP80944.
XX
PT   Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
XX
PS   Disclosure; ; pp; English.
XX
CC   The sequence can be cloned and used to produce recombinant LEP-10.
CC   The DNA can be used as a probe for identifying further coding
CC   sequences for protease inhibitors; the protein is useful as an
CC   assay reagent.
CC   See also AAN80926.
XX
SQ   Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 9;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatta 22
          ||||||||||||||
Db    251 AAGTCACAGAATTA 238


RESULT  17
AAT31932/c
ID   AAT31932 standard; DNA; 1677 BP.
XX
AC   AAT31932;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-2 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-2; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1356
FT                   /*tag=  a
FT   polyA_signal    1672-1677
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99739.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 59-60; 90pp; English.
XX
CC   A cDNA clone (AAT31932) codes for retinoid X receptor interacting
CC   protein RIP14-2 (AAR99739), a previously undescribed orphan member
CC   of the nuclear receptor superfamily which may be involved in the
CC   complex retinoid response.  It was isolated from a mouse liver
CC   cDNA library using an RIP14 probe.  Different RIP14-2 clones had
CC   different 5' sequences (see also AAT31935-37).  Another isoform,
CC   RIP14-1, was also detected (see also AAT31928).  The cDNA can be
CC   used for the recombinant prodn. of RIP14-2 in transformed host
CC   cells.
XX
SQ   Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    630 TTGTCGCAAGTCAC 617


RESULT  18
AAT31928/c
ID   AAT31928 standard; DNA; 1787 BP.
XX
AC   AAT31928;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-1 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..1466
FT                   /*tag=  a
FT   polyA_signal    1782..1787
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99735.
```

```
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 53-54; 90pp; English.
XX
CC   A full-length cDNA clone (AAT31928) codes for retinoid X receptor
CC   interacting protein RIP14-1 (AAR99735), a previously undescribed
CC   orphan member of the nuclear receptor superfamily which may be
CC   involved in the complex retinoid response.  It was isolated from
CC   a mouse liver cDNA library using an RIP14 probe.  Another isoform,
CC   RIP14-2, was also detected (see also AAT31932).  The cDNA can be used
CC   for the recombinant prodn. of RIP14-1 in transformed host cells.
XX
SQ   Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    740 TTGTCGCAAGTCAC 727


RESULT  19
AAZ52940/c
ID   AAZ52940 standard; cDNA; 1890 BP.
XX
AC   AAZ52940;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #83.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; AAY74061, AAY74062, AAY74063.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 2; Page 250; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1890;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatt 21
          ||||||||||||||
Db   1475 CAAGTCACAGAATT 1462


RESULT  20
AAV10120/c
ID   AAV10120 standard; cDNA; 2233 BP.
XX
AC   AAV10120;
XX
DT   29-MAY-1998  (first entry)
XX
DE   Human retinoid receptor RR1 cDNA.
XX
KW   Retinoid receptor; RR1; steroid receptor; agonist; antagonist; cancer;
KW   adrenal deficiency; skin disorder; inflammatory disorder;
KW   immune response regulator; autoimmune disease; therapeutic antibody; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             363..1781
FT                   /*tag=  a
FT                   /product= RR1
FT                   /note= "retinoid receptor"
XX
PN   US5728548-A.
XX
PD   17-MAR-1998.
XX
PF   29-JUN-1995;   95US-0496631.
XX
PR   29-JUN-1995;   95US-0496631.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Bowman M;
XX
DR   WPI; 1998-206567/18.
DR   P-PSDB; AAW40072.
XX
PT   Human retinoid receptor protein RR1 - useful for, e.g. drug
PT   screening, therapy and antibody production
XX
PS   Claim 8; Column 11-15; 13pp; English.
XX
CC   This sequence encodes a novel human steroid receptor, the retinoid
CC   receptor protein or RR1. This protein can be used in screening assays
CC   for steroid hormone receptor agonists and antagonists and in
CC   pharmaceutical compositions for treating adrenal deficiencies, e.g.
CC   Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
CC   inflammatory disorders, e.g. arthritis and HIV infections. The protein
CC   can also be used for regulating immune responses, e.g. as antitumour
CC   agents, vaccine adjuvants, organ rejection inhibitors or agents for
CC   treating autoimmune diseases. The protein can further be used to produce
CC   therapeutic antibodies.
XX
SQ   Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 19;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db   1055 TTGTCGCAAGTCAC 1042
```

```
RESULT  21
AAF77541/c
ID   AAF77541 standard; DNA; 2585 BP.
XX
AC   AAF77541;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Caenorhabditis briggsae mab-21 coding sequence.
XX
KW   Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
KW   chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
OS   Caenorhabditis briggsae.
XX
PN   US6197504-B1.
XX
PD   06-MAR-2001.
XX
PF   19-JAN-1998;   98US-0008697.
XX
PR   10-APR-1996;   96US-0011607.
PR   09-APR-1997;   97US-0835604.
XX
PA   (CHOW/) CHOW K L.
XX
PI   Chow KL;
XX
DR   WPI; 2001-256362/26.
DR   P-PSDB; AAB74419.
XX
PT   Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
PT   novel protein required for choice of alternate cell fates in sample, by
PT   contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
PS   Disclosure; column 41-44; 79pp; English.
XX
CC   The present invention describes a method of detecting mab-21 expression
CC   in a sample involving contacting total mRNA from the sample with a probe
CC   to the mab-21 gene. The sequences of the mab-21 coding sequence and
CC   protein from several species are also given. The human mab-21 gene is
CC   found on chromosome 13q13, a region which has been linked to Moebius
CC   syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC   are associated with aniridia and which is linked to peripheral neural
CC   tissue differentiation.
XX
SQ   Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatt 21
          ||||||||||||||
Db   1058 CAAGTCACAGAATT 1045


RESULT  22
AAC69462/c
ID   AAC69462 standard; cDNA; 2845 BP.
XX
AC   AAC69462;
XX
DT   30-JAN-2001  (first entry)
XX
DE   Human secreted protein gene 8 SEQ ID NO:18.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; skin aging; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058469-A1.
XX
PD   05-OCT-2000.
XX
PF   23-MAR-2000; 2000WO-US07579.
XX
PR   26-MAR-1999;   99US-0126509.
PR   07-JAN-2000; 2000US-0174853.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594642/56.
DR   P-PSDB; AAB38210.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 334-335; 416pp; English.
XX
CC   The polynucleotide sequences given in AAC69455 to AAC69502 encode the
CC   human secreted proteins given in AAB38203 to AAB38250. AAB38251 to
CC   AAB38320 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Example of activities include:
CC   immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC   neuroprotective; antibacterial; virucide; fungicide; and
CC   ophthalmological. The polynucleotides and polypeptides can be used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. AAC69446 to AAC69454 and
CC   AAB38202 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2845 BP; 871 A; 518 C; 439 G; 1007 T; 10 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 2845;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 tcacagaattattt 25
          ||||||||||||||
Db   1681 TCACAGAATTATTT 1668


RESULT  23
AAV72330/c
ID   AAV72330 standard; DNA; 3413 BP.
XX
AC   AAV72330;
XX
DT   28-JUL-1999  (first entry)
XX
```

DE F. balustinum CP70 DNA.
XX
KW CP70; cold active protease; detergent; cold temperature wash water; ss.
XX
OS Flavobacterium balustinum.
XX
FH Key Location/Qualifiers
FT CDS 437..2536
FT /*tag= a
FT /product= "CP70"
XX
PN WO9925848-A1.
XX
PD 27-MAY-1999.
XX
PF 14-NOV-1997; 97WO-US20843.
XX
PR 14-NOV-1997; 97WO-US20843.
XX
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Kitado H, Yoshikawa A;
XX
DR WPI; 1999-347483/29.
DR P-PSDB; AAY08471.
XX
PT Newly isolated CP70 polynucleotide exhibiting characteristics of
PT Deposit No. FERM BP-6154
XX
PS Claim 1a; Page 15-18; 25pp; English.
XX
CC This invention describes a novel cold active protease, CP70 from
CC Flavobacterium balustinum. The product of the invention is useful
CC in detergent compositions for use in cold temperature wash water.
CC The characterization of the polynucleotide facilitates the production
CC of an alternative relatively large source of cold active protease.
XX
SQ Sequence 3413 BP; 1228 A; 558 C; 586 G; 1041 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 3413;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattat 23
||||||||||||||
Db 3107 AGTCACAGAATTAT 3094

RESULT 24
AAC59953
ID AAC59953 standard; cDNA; 5684 BP.
XX
AC AAC59953;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #47.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200055198-A1.
XX
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06012.
XX
PR 12-MAR-1999; 99US-0124093.
PR 23-NOV-1999; 99US-0166989.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-587520/55.
DR P-PSDB; AAB28748.
XX
PT Human secreted proteins and the nucleic acids that encode them, useful
PT in gene therapy protocols and recombinant nucleic acid based procedures
PT -
XX
PS Claim 1; Page 343-344; 391pp; English.
XX
CC The invention relate to the isolation of genes AAC59907-C59956 encoding
CC 50 human secreted proteins AAB28702-B28751. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 5684 BP; 1641 A; 1250 C; 1045 G; 1732 T; 16 other;

Query Match 56.0%; Score 14; DB 21; Length 5684;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattattt 25
|||||||||||||||
Db 4933 tcacagaattattt 4946

RESULT 25
AAC76014/c
ID AAC76014 standard; cDNA; 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX

OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41805.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2354-2357; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 5773;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       7 gcaagtcacagaat 20
           ||||||||||||||
Db    4611 GCAAGTCACAGAAT 4598
```

RESULT 26
AAT10105
ID AAT10105 standard; DNA; 8041 BP.
XX
AC AAT10105;
XX
DT 13-MAY-1996 (first entry)
XX
DE Adherence conferring plasmid pear.
XX
KW adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.
XX
OS Chimeric Escherichia coli;

OS Chimeric synthetic.
XX
FH Key Location/Qualifiers
FT primer_bind complement (2867..2890)
FT /*tag= a
FT /note= "primer for E.coli O157:H7 DNA"
FT primer_bind complement (2895..2914)
FT /*tag= b
FT /note= "primer for pear"
FT CDS 3036..5126
FT /*tag= c
FT /product= adhesin
FT /trans_except=pos:3912..3914:_aa:Gly
FT /note= "claim 1, page 36"
FT misc_feature 3271..3310
FT /*tag= d
FT /note= "region of transposon TnphoA insertion"
FT misc_feature 3801..3840
FT /*tag= e
FT /note= "region of transposon TnPhoA insertion"
FT primer_bind complement (5176..5196)
FT /*tag= f
FT /note= "primer for pear"
FT primer_bind complement (5159..5183)
FT /*tag= g
FT /note= "primer for E.coli O157H7 DNA"
FT CDS complement (6449..7024)
FT /*tag= h
FT /note= "terE homologue"
FT CDS complement (7092..7670)
FT /*tag= i
FT /note= "terD homologue"
XX
PN WO9600233-A1.
XX
PD 04-JAN-1996.
XX
PF 07-JUN-1995; 95WO-US06994.
XX
PR 24-JUN-1994; 94US-0265714.
XX
PA (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA (UNIW) UNIV WASHINGTON.
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI Besser TE, Bilge SS, Tarr PI, Vary JC;
XX
DR WPI; 1996-068826/07.
DR P-PSDB; AAR75366.
XX
PT Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT isolated on plasmid pSC (overlap), for use as a vaccine to prevent
PT bacterial colonisation of bovine intestine
XX
PS Disclosure; Page 23-30; 42pp; English.
XX
CC Adherence conferring plasmid pear (AAT10105) comprises Escherichia
CC coli O157:H7 chromosomal DNA plus the Stratagene SK+ vector.
CC It includes 3 open reading frames, 2 of which are homologues of
CC terE and terD genes necessary for tellurite resistance. The
CC third ORF is homologous to the IrgA gene and encodes an adhesin
CC (AAR75366) that enables E. coli O157:H7, an antibiotic-resistant,
CC virulent and common food-borne pathogen, to adhere to epithelial
CC cells. This ORF can be utilised in the prodn. of adhesin for use
CC as a vaccine to prevent disease or colonisation of mucosal surfaces
CC by O157:H7.
XX
SQ Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 29 other;

Query Match 56.0%; Score 14; DB 17; Length 8041;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 agtcacagaattat 3740


RESULT  27
AAF21307
ID   AAF21307 standard; DNA; 32351 BP.
XX
AC   AAF21307;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2874.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1295-1303; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 32351;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatta 22
           ||||||||||||||
Db   10696 aagtcacagaatta 10709


RESULT  28
AAA35185
ID   AAA35185 standard; DNA; 32351 BP.
XX
AC   AAA35185;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1212-1219; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
```

CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX
SQ Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatta 22
||||||||||||||
Db 10696 aagtcacagaatta 10709

RESULT 29
AAF21311
ID AAF21311 standard; DNA; 40298 BP.
XX
AC AAF21311;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2878.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1305-1315; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 40298;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatta 22
||||||||||||||
Db 11521 aagtcacagaatta 11534

RESULT 30
AAA35189
ID AAA35189 standard; DNA; 40298 BP.
XX
AC AAA35189;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US17712.
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.

```
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1221-1231; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 40298;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatta 22
           ||||||||||||||
Db   11521 aagtcacagaatta 11534


RESULT  31
AAF30035
ID   AAF30035 standard; DNA; 81001 BP.
XX
AC   AAF30035;
XX
DT   23-APR-2001  (first entry)
XX
DE   Human apolipoprotein A-IV-related protein (AA4RP) gene.
XX
KW   Apolipoprotein A-IV-related protein; AA4RP; human;
KW   biallelic marker; lipid metabolism; liver related disorder;
KW   obesity; diabetes; coronary heart disease; diagnosis; gene therapy;
KW   chromosome 11; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           10946..12946
FT                   /*tag=  a
FT   CDS             13478..15210
FT                   /*tag=  b
FT   exon            12947..12958
FT                   /*tag=  c
FT                   /number= 1
FT   intron          12959..13469
FT                   /*tag=  d
FT                   /number= 1
FT   exon            13470..13526
FT                   /*tag=  e
FT                   /number= 2
FT   intron          13527..13640
FT                   /*tag=  f
FT                   /number= 2
FT   exon            13641..13752
FT                   /*tag=  g
FT                   /number= 3
FT   intron          13753..14270
FT                   /*tag=  h
FT                   /number= 3
FT   exon            14271..15968
FT                   /*tag=  i
FT                   /number= 4
FT   3'UTR           15969..17969
FT                   /*tag=  j
FT   allele          replace(1239,C,T)
FT                   /*tag=  k
FT                   /note= "biallelic marker 20-828-311, polymorphic
FT                          base C or T"
FT   allele          replace(12347,C,T)
FT                   /*tag=  l
FT                   /note= "biallelic marker 17-42-319, polymorphic
FT                          base C or T"
FT   allele          replace(15241,C,T)
FT                   /*tag=  m
FT                   /note= "biallelic marker 17-41-250, polymorphic
FT                          base C or T"
FT   allele          replace(42218,A,G)
FT                   /*tag=  n
FT                   /note= "biallelic marker 20-841-149, polymorphic
FT                          base A or G"
FT   allele          replace(45442,A,G)
FT                   /*tag=  o
FT                   /note= "biallelic marker 20-842-115, polymorphic
FT                          base A or G"
FT   allele          replace(77058,C,T)
FT                   /*tag=  p
FT                   /note= "biallelic marker 20-853-415, polymorphic
FT                          base C or T"
FT   primer_bind     929..949
FT                   /*tag=  q
FT                   /note= "primer 20-828.pu"
FT   primer_bind     complement (1357..1377)
FT                   /*tag=  r
FT                   /note= "primer 20-828.rp"
FT   primer_bind     12029..12050
FT                   /*tag=  s
FT                   /note= "primer 17-42.pu"
FT   primer_bind     complement (12581..12603)
FT                   /*tag=  t
FT                   /note= "primer 17-42.rp"
FT   primer_bind     14992..15012
FT                   /*tag=  u
FT                   /note= "primer 17-41.pu"
FT   primer_bind     complement (15460..15482)
FT                   /*tag=  v
FT                   /note= "primer 17-41.rp"
FT   primer_bind     42070..42090
FT                   /*tag=  w
FT                   /note= "primer 20-841.pu"
FT   primer_bind     complement (42572..42591)
FT                   /*tag=  x
FT                   /note= "primer 20-841.rp"
FT   primer_bind     45328..45347
FT                   /*tag=  y
FT                   /note= "primer 20-842.pu"
FT   primer_bind     complement (45863..45883)
FT                   /*tag=  z
FT                   /note= "primer 20-842.rp"
FT   primer_bind     76644..76664
```

```
FT                   /*tag=  aa
FT                   /note= "primer 20-853.pu"
FT   primer_bind     complement (77166..77185)
FT                   /*tag=  ab
FT                   /note= "primer 20-842.rp"
FT   primer_bind     1220..1238
FT                   /*tag=  ac
FT                   /note= "primer 20-828-311.mis"
FT   primer_bind     complement (1240..1258)
FT                   /*tag=  ad
FT                   /note= "primer 20-828-311.mis"
FT   primer_bind     12328..12346
FT                   /*tag=  ae
FT                   /note= "primer 17-42-319.mis"
FT   primer_bind     complement (12348..12366)
FT                   /*tag=  af
FT                   /note= "primer 17-42-319.mis"
FT   primer_bind     15222...15240
FT                   /*tag=  ag
FT                   /note= "primer 17-41-250.mis"
FT   primer_bind     complement (15242..15260)
FT                   /*tag=  ah
FT                   /note= "primer 17-41-250.mis"
FT   primer_bind     42199..42237
FT                   /*tag=  ai
FT                   /note= "primer 20-841-149.mis"
FT   primer_bind     complement (42219..42237)
FT                   /*tag=  aj
FT                   /note= "primer 20-841-149.mis"
FT   primer_bind     45423..45441
FT                   /*tag=  ak
FT                   /note= "primer 20-842-115.mis"
FT   primer_bind     complement (45443..45461)
FT                   /*tag=  al
FT                   /note= "primer 20-842-115.mis"
FT   primer_bind     77039..77057
FT                   /*tag=  am
FT                   /note= "primer 20-853-415.mis"
FT   primer_bind     complement (77059..77077)
FT                   /*tag=  an
FT                   /note= "primer 20-853-415.mis"
FT   misc_binding    1227..1251
FT                   /*tag=  ao
FT                   /note= "20-828-311 probe"
FT   misc_binding    12335..12359
FT                   /*tag=  ap
FT                   /note= "17-42-319 probe"
FT   misc_binding    15229..15253
FT                   /*tag=  aq
FT                   /note= "17-41-250 probe"
FT   misc_binding    42206..42230
FT                   /*tag=  ar
FT                   /note= "20-841-149 probe"
FT   misc_binding    45430..45454
FT                   /*tag=  as
FT                   /note= "20-842-115 probe"
FT   misc_binding    77046..77070
FT                   /*tag=  at
FT                   /note= "20-853-415 probe"
XX
PN   WO200100803-A2.
XX
PD   04-JAN-2001.
XX
PF   21-JUN-2000; 2000WO-IB01011.
XX
PR   25-JUN-1999;   99US-0141032.
PR   20-DEC-1999;   99WO-IB02058.
PR   21-DEC-1999;   99US-0469099.
XX
PA   (GEST ) GENSET.
XX
PI   Yen F,  Denison B,  Bour B,  Bihain B,  Bougueleret L,  Duclert A;
PI   Dumas Milne Edwards J;
XX
DR   WPI; 2001-071485/08.
DR   P-PSDB; AAB20103.
XX
PT   Nucleic acids encoding apolipoprotein A-IV-related proteins (AA4RP) and
PT   biallelic markers of AA4RP, useful for diagnosing lipid metabolism
PT   related disorders and/or liver related disorders -
XX
PS   Claim 1; Page 216-238; 260pp; English.
XX
CC   The present sequence is that of the human apolipoprotein
CC   A-IV-related protein (AA4RP) gene on chromosome 11.  The gene
CC   encodes a 366-amino acid protein (see AAB20103).  AA4RP is
CC   differentially expressed in obese mouse models, indicating a role
CC   in lipid metabolism related disorders.  It appears to be the human
CC   homologue of rat regeneration associated protein (RAP3), which is
CC   believed to be involved in liver regeneration, and is likely to
CC   have a similar function.  It also shows 52% similarity to
CC   apolipoprotein A-IV, and is likely to have a similar function.  The
CC   invention also provides AA4RP cDNAs (see AAF30036) and polypeptides,
CC   biallelic markers identified in the AA4RP gene and from genomic
CC   regions flanking the gene, and methods for genotyping a nucleic
CC   acid containing 1 or more of the biallelic markers.  Also provided
CC   is a method for detecting a statistical correlation between a
CC   biallelic marker allele and a phenotype and/or between a biallelic
CC   marker haplotype and a phenotype.  Diagnostic methods are provided
CC   for determining whether an individual is at risk of developing a
CC   lipid metabolism related disorder and/or a liver related disorder,

  Query Match             56.0%;  Score 14;  DB 22;  Length 81001;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattatt 24
           ||||||||||||||
Db   58568 gtcacagaattatt 58581


RESULT  32
AAZ01425
ID   AAZ01425 standard; DNA; 1038602 BP.
XX
AC   AAZ01425;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Complete genome sequence of Chlamydia trachomatis.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
```

```
XX
PS   Claim 1; Page 373-656; 1755pp; English.
XX
CC   The present sequence represents the complete genome of Chlamydia
CC   trachomatis. Open reading frames (ORFs) of the genome encode
CC   polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC   against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC   be used to control growth of the microorganism. Chlamydia trachomatis is
CC   responsible for a large number of diseases, e.g. eye diseases such as
CC   conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC   conjunctivitis; genital diseases such as nongonococcal uretritis,
CC   epidymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
CC   pneumopathy in breast feeding infants; and venereal
CC   lymphogranulomatosis. The polypeptides of the invention may be of use in
CC   treating these diseases.
XX
SQ   Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1038602;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtca 14
          ||||||||||||||
Db 736214 attgtcgcaagtca 736227


RESULT  33
AAT22576
ID   AAT22576 standard; cDNA to mRNA; 87 BP.
XX
AC   AAT22576;
XX
DT   01-OCT-1996  (first entry)
XX
DE   Human gene signature HUMGS04192.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1160; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;


  Query Match             52.0%;  Score 13;  DB 16;  Length 87;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatt 21
          |||||||||||||
Db     16 aagtcacagaatt 28


RESULT  34
AAT20380
ID   AAT20380 standard; cDNA to mRNA; 105 BP.
XX
AC   AAT20380;
XX
DT   19-JUL-1996  (first entry)
XX
DE   Human gene signature HUMGS01532.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 625; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
```

```
CC   recognising different cell types.
XX
SQ   Sequence 105 BP; 26 A; 19 C; 20 G; 34 T; 6 other;


  Query Match             52.0%;  Score 13;  DB 16;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 cacagaattattt 25
           |||||||||||||
Db      92 cacagaattattt 104


RESULT  35
AAT20233/c
ID   AAT20233 standard; cDNA to mRNA; 167 BP.
XX
AC   AAT20233;
XX
DT   24-JUL-1996  (first entry)
XX
DE   Human gene signature HUMGS01380.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 592; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;


  Query Match             52.0%;  Score 13;  DB 16;  Length 167;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaatta 22
           |||||||||||||
Db     118 AGTCACAGAATTA 106


RESULT  36
AAC26213/c
ID   AAC26213 standard; cDNA; 187 BP.
XX
AC   AAC26213;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30288.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30288; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 187;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     163 AAGTCACAGAATT 151


RESULT  37
AAQ77082
ID   AAQ77082 standard; DNA; 354 BP.
XX
```

```
AC   AAQ77082;
XX
DT   23-SEP-1994  (first entry)
XX
DE   Human genome fragment.
XX
KW   Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW   detection; homology; human; adrenal tissue; ds.
XX
OS   Homo sapiens.
XX
PN   WO9401548-A.
XX
PD   20-JAN-1994.
XX
PF   13-JUL-1993;   93WO-GB01467.
XX
PR   13-JUL-1992;   92GB-0014857.
XX
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Gross J,  Hadfield KM,  Howells D,  Kelly M,  Shaw D;
PI   Sibson DR,  Starkey M;
XX
DR   WPI; 1994-035056/04.
XX
PT   New nucleic acid fragment encoding gene products - can be used
PT   for genetic analysis and mapping
XX
PS   Claim 1; Page 347; 616pp; English.
XX
CC   Human nucleic acid fragments, isolated from brain adrenal tissue,
CC   the placenta or bone marrow comprise any of: (A) a sequence
CC   selected from (AAQ76401-Q77613), (B) an allelic variation of a
CC   sequence as described in (A), or (C) a sequence complementary
CC   to (A) or (B).
XX
SQ   Sequence 354 BP; 92 A; 92 C; 86 G; 84 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 15;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     192 tcacagaattatt 204


RESULT  38
AAA50159/c
ID   AAA50159 standard; cDNA; 363 BP.
XX
AC   AAA50159;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Antibody 5H7 heavy chain variable region cDNA.
XX
KW   5H7; monoclonal antibody; apoptosis; programmed cell death;
KW   major histocompatibility complex class I; MHC; tumour; therapy;
KW   graft-mediated immunoprotection; single chain antibody; scFv; ss.
XX
OS   Mammalia.
XX
PN   WO200047713-A2.
XX
PD   17-AUG-2000.
XX
PF   08-FEB-2000; 2000WO-US03234.
XX
PR   09-FEB-1999;   99US-0119238.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Woodle ES,  Van Seventer JM,  Kulkarni S,  Kranz D,  Holman P;
XX
DR   WPI; 2000-558210/51.
DR   P-PSDB; AAY95776.
XX
PT   New polynucleotides encoding single chain light and heavy variable
PT   regions of a 5H7 antibody, useful in methods for conferring programmed
PT   cell death and graft-mediated immune protection of cell, tissue and
PT   organs -
XX
PS   Claim 2; Page 36; 48pp; English.
XX
CC   The present sequence is that of cDNA coding for the heavy chain
CC   variable region (VH) (see AAY95776) of 5H7, an anti-human class I
CC   major histocompatibility complex (MHC) monoclonal antibody (MAb)
CC   which recognises a monomorphic determinant of the alpha-3 domain.
CC   MAb 5H7 has been shown to induce programmed cell death (PCD) in
CC   lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
CC   splenocytes transgenic for class I HLA-B27.  Polynucleotides
CC   encoding 5H7 VL and VH sequences were used in the construction of
CC   single chain variable immunoglobulin domains (scFv) of the 5H7
CC   antibody (see AAY95781).  These can be used to confer PCD properties
CC   to cells, especially T and B lymphocyte tumour cells.  Recombinant
CC   DNA molecules encoding 5H7 scFv are also used in methods for
CC   conferring graft-mediated immune protection of cell, tissue and
CC   transplanted organs.
XX
SQ   Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     262 AAGTCACAGAATT 250


RESULT  39
AAQ25667/c
ID   AAQ25667 standard; cDNA; 378 BP.
XX
AC   AAQ25667;
XX
DT   28-DEC-1992  (first entry)
XX
DE   Sequence of the anti-urokinase antibody kappa variable region (VK)
DE   cDNA.
XX
KW   Chimeric monoclonal antibody; anti-urokinase antibody; PCR;
KW   antithrombotic agent; myocardial infarction therapy; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..9
FT                   /*tag=  a
FT   mat_peptide     10..378
FT                   /*tag=  b
FT   misc_feature    1..27
FT                   /*tag=  c
FT                   /label= VH01 primer
FT   misc_feature    18..32
FT                   /*tag=  d
FT                   /label= 5'VH1 primer
FT   misc_feature    340..371
FT                   /*tag=  e
FT                   /label= 3'mVH primer
FT   misc_feature    356..371
FT                   /*tag=  f
```

```
FT                  /label= JH01 primer
XX
PN   EP491351-A.
XX
PD   24-JUN-1992.
XX
PF   17-DEC-1991;   91EP-0121591.
XX
PR   18-DEC-1990;   90JP-0413829.
PR   11-NOV-1991;   91JP-0294464.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Iwasa S,  Taka H,  Watanabe T,  Tada H;
XX
DR   WPI; 1992-209528/26.
DR   P-PSDB; AAR24722.
XX
PT   Chimeric monoclonal antibodies - contain anti-human fibrin
PT   antibody light and heavy chain variable and constant for treating
PT   thrombotic conditions e.g. myocardial infarction
XX
PS   Example; Figure 15; 87pp; English.
XX
CC   Poly(A)+ RNA was prepd. from mouse anti-urokinase
CC   antibody-producing hybridoma UK1-3 cells. Using this poly(A) RNA as
CC   a template, an anti-urokinase antibody VK cDNA was amplified with
CC   the mC-kappa primer as primer for first strand sysnthesis, and the
CC   3'mV-kappa and 5'mV-kappa primers as primers for the PCR. The
CC   amplified fragment was restriction digested and ligated into a
CC   restriction fragment of pTB1423 to give an anti-urokinase antibody
CC   VK cDNA contg. plasmid pTB1456. The sequence of this plasmid is
CC   given in AAQ25667. The cDNA is a functional VK gene. The sequence of
CC   the primer (mC-gamma-1) is given in AAQ25689.
XX
SQ   Sequence 378 BP; 92 A; 100 C; 91 G; 95 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 13;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     271 AAGTCACAGAATT 259


RESULT  40
AAC79285
ID   AAC79285 standard; cDNA; 387 BP.
XX
AC   AAC79285;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Human lung tumour-specific cDNA #238.
XX
KW   Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200060077-A2.
XX
PD   12-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08560.
XX
PR   02-APR-1999;   99US-0285323.
PR   09-AUG-1999;   99US-0370838.
PR   30-DEC-1999;   99US-0476235.
PR   03-MAR-2000; 2000US-0518809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Lodes MJ,  Mohamath R,  Secrist H;
XX
DR   WPI; 2000-638466/61.
XX
PT   Novel lung tumor polypeptides and polynucleotides, useful for
PT   detecting, monitoring or treating cancer, especially lung cancer -
XX
PS   Claim 24; Page 223; 243pp; English.
XX
CC   The present sequence is given in a specification relating to compounds
CC   for therapy and diagnosis of lung cancer. Polypeptides comprising at
CC   least an immunogenic part of a lung tumour protein are disclosed.
CC   The polypeptides are useful for inhibiting the development of cancer,
CC   especially lung cancer. Samples of T cells expressing the polypeptides
CC   may be used to inhibit the development of cancer. The polypeptides are
CC   also useful for detecting and monitoring the progression of cancer,
CC   especially lung cancer.
XX
SQ   Sequence 387 BP; 95 A; 91 C; 99 G; 102 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     179 tcacagaattatt 191


RESULT  41
AAA74604/c
ID   AAA74604 standard; cDNA; 408 BP.
XX
AC   AAA74604;
XX
DT   08-JAN-2001  (first entry)
XX
DE   Murine 5B3 antibody heavy chain variable region cDNA.
XX
KW   Mouse; 5B3 antibody; IgG1; piezoelectric immunoassay;
KW   small molecule explosive detection; 2,4,6-trinitrotoluene; TNT; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..408
FT                   /*tag=  a
FT                   /partial
FT                   /product= "5B3 antibody heavy chain variable region"
XX
PN   WO200043774-A2.
XX
PD   27-JUL-2000.
XX
PF   25-JAN-2000; 2000WO-IL00048.
XX
PR   25-JAN-1999;   99IL-0128212.
XX
PA   (YISS ) YISSUM RES & DEV CO.
PA   (YEDA ) YEDA RES & DEV CO LTD.
XX
PI   Willner I,  Eshhar Z;
XX
DR   WPI; 2000-524259/47.
DR   P-PSDB; AAB15672.
XX
PT   Apparatus for detecting small molecules, especially explosives
PT   comprises a piezoelectric crystal -
XX
PS   Disclosure; Fig 3A; 90pp; English.
XX
```

```
CC   The present sequence encodes the heavy chain variable region of 5B3
CC   antibody. It was obtained from total RNA extracted from a 5B3 hybridoma
CC   by RT-PCR. 5B3 is an IgG1 antibody derived from a TNP-KLH immunised
CC   mouse. Its binding to TNP or DNS antigen can be blocked by very low
CC   amounts of TNT and it can therefore be used in a method for detecting
CC   small assayed explosive molecules. Molecules are detected using a
CC   piezoelectric sensor. Piezoelectric immunoassaying in liquid phase
CC   allows stationary and flow analysis of an aqueous sample. The method
CC   is sufficiently sensitive for detection of low molecular weight
CC   molecules.
XX
SQ   Sequence 408 BP; 90 A; 108 C; 99 G; 111 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     316 AAGTCACAGAATT 304


RESULT  42
AAC03836
ID   AAC03836 standard; cDNA; 409 BP.
XX
AC   AAC03836;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 3834.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG03830.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 3834; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 409 BP; 95 A; 101 C; 118 G; 91 T; 4 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     233 tcacagaattatt 245


RESULT  43
AAC56783/c
ID   AAC56783 standard; DNA; 415 BP.
XX
AC   AAC56783;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #654.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Pages 513-514; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 415;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 attgtcgcaagtc 13
          |||||||||||||
Db    121 ATTGTCGCAAGTC 109


RESULT  44
AAC07649/c
ID   AAC07649 standard; cDNA; 441 BP.
XX
AC   AAC07649;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 11724.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 11724; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 441 BP; 123 A; 103 C; 113 G; 102 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 cacagaattattt 25
           |||||||||||||
Db     419 CACAGAATTATTT 407


RESULT  45
AAC07951
ID   AAC07951 standard; cDNA; 476 BP.
XX
AC   AAC07951;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 12026.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 12026; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 476 BP; 117 A; 113 C; 135 G; 107 T; 4 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     300 tcacagaattatt 312


Search completed: October  9, 2001, 15:52:06
Job time: 15087 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:17 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8765

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c 5	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli
6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appl
c 7	14	56.0	1050	6	5498529-7	Patent No. 5498529
c 8	14	56.0	1677	2	US-08-372-652-14	Sequence 14, Appl
c 9	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appl
c 10	14	56.0	1787	2	US-08-372-652-6	Sequence 6, Appli
c 11	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appli
12	14	56.0	2091	1	US-08-765-081-4	Sequence 4, Appli
13	14	56.0	2091	3	US-09-098-082-4	Sequence 4, Appli
14	14	56.0	2091	5	PCT-US95-06994-4	Sequence 4, Appli
c 15	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appli
c 16	14	56.0	2585	4	US-09-008-697A-7	Sequence 7, Appli
17	14	56.0	8041	1	US-08-765-081-1	Sequence 1, Appli
18	14	56.0	8041	3	US-09-098-082-1	Sequence 1, Appli
19	14	56.0	8041	5	PCT-US95-06994-1	Sequence 1, Appli
c 20	13	52.0	1500	4	US-08-875-847B-3	Sequence 3, Appli
c 21	13	52.0	1505	4	US-08-875-847B-1	Sequence 1, Appli
c 22	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appli
c 23	13	52.0	1662	3	US-09-066-544-1	Sequence 1, Appli
c 24	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appli
c 25	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appli
c 26	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appli
c 27	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appli
c 28	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appli
c 29	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appli
c 30	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appli
c 31	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appli
c 32	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appli
c 33	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appli
c 34	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appli
c 35	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appli
c 36	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appli
c 37	13	52.0	3083	1	US-07-718-575-5	Sequence 5, Appli
c 38	13	52.0	3083	1	US-08-481-206-5	Sequence 5, Appli
c 39	13	52.0	3083	2	US-08-486-269A-5	Sequence 5, Appli
c 40	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
c 41	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
42	13	52.0	5590	4	US-09-050-159-129	Sequence 129, App
c 43	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appli
44	12	48.0	17	1	US-08-373-124A-284	Sequence 284, App
45	12	48.0	17	1	US-08-373-124A-286	Sequence 286, App

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7
```

```
  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match           100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
```

```
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792
;  GENERAL INFORMATION:
;    APPLICANT:  Ifuku, Ohji
;    APPLICANT:  Haze, Shinitiro
;    APPLICANT:  Kishimoto, Jiro
;    APPLICANT:  Nakahama, Kazuo
;    TITLE OF INVENTION:  BIOTIN OPERON
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wegner, Cantor, Mueller & Player
;      STREET:  1233 20th Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-8218
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/944,259
;      FILING DATE:  19920914
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cantor, Herbert I.
;      REGISTRATION NUMBER:  24,392
;      REFERENCE/DOCKET NUMBER:  P-450-23557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-0400
;      TELEFAX:  202-835-0605
;      TELEX:  440706
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
US-07-944-259-2


  Query Match             68.0%;  Score 17;  DB 2;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gtcgcaagtcacagaat 20
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1


RESULT   6
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0323
;  CURRENT APPLICATION NUMBER: US/09/166,203A
;  CURRENT FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 60
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48


  Query Match             60.0%;  Score 15;  DB 2;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   7
5498529-7/c
;Patent No. 5498529
;     APPLICANT: BERKA, THOMAS R.;FORNWALD, JAMES A.;GORNIAK,
;JOSELINA G.;ROSENBERG, MARTIN;STRICKLER, JAMES E.;TAYLOR, DEAN P.
;     TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
;STREPTOMYCES
;     NUMBER OF SEQUENCES: 9
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,506
;       FILING DATE: 03-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 894,167
;       FILING DATE: 03-JUN-1992
;       APPLICATION NUMBER: 346,119
;       FILING DATE: 16-FEB-1989
;       APPLICATION NUMBER: 897,245
;       FILING DATE: 18-AUG-1986
;SEQ ID NO:7:
;       LENGTH: 1050
5498529-7


  Query Match             56.0%;  Score 14;  DB 6;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatta 22
           ||||||||||||||
Db     251 AAGTCACAGAATTA 238


RESULT   8
US-08-372-652-14/c
; Sequence 14, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
```

```
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1677 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-372-652-14


  Query Match             56.0%;  Score 14;  DB 2;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    630 TTGTCGCAAGTCAC 617


RESULT  9
PCT-US95-16311-14/c
; Sequence 14, Application PC/TUS9516311
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16311
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1677 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16311-14


  Query Match             56.0%;  Score 14;  DB 5;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    630 TTGTCGCAAGTCAC 617


RESULT  10
US-08-372-652-6/c
; Sequence 6, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1787 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-372-652-6
```

```
  Query Match             56.0%;  Score 14;  DB 2;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    740 TTGTCGCAAGTCAC 727


RESULT  11
PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16311
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1787 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
PCT-US95-16311-6


  Query Match             56.0%;  Score 14;  DB 5;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    740 TTGTCGCAAGTCAC 727


RESULT  12
US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-08-765-081-4


  Query Match             56.0%;  Score 14;  DB 1;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    692 AGTCACAGAATTAT 705


RESULT  13
US-09-098-082-4
; Sequence 4, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
```

```
;       OPERATING SYSTEM:  MS-Windows 95
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/098,082
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/765,081
;       FILING DATE:  March 26, 1997
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sheiness, Diana K.
;       REGISTRATION NUMBER:  35,356
;       REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2091 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;       DESCRIPTION:  Corresponds to SEQ ID NO:1,
;       DESCRIPTION:  nucleotides 3036-5126
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157: H7
;       STRAIN:  86-24 NALR
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..2088
US-09-098-082-4


  Query Match             56.0%;  Score 14;  DB 3;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db     692 AGTCACAGAATTAT 705


RESULT  14
PCT-US95-06994-4
; Sequence 4, Application PC/TUS9506994
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
PCT-US95-06994-4


  Query Match             56.0%;  Score 14;  DB 5;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db     692 AGTCACAGAATTAT 705


RESULT  15
US-08-496-631-1/c
; Sequence 1, Application US/08496631
; Patent No. 5728548
;  GENERAL INFORMATION:
;    APPLICANT:  Bowman, Michael
;    TITLE OF INVENTION:  STEROID RECEPTOR RR1
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/496,631
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
```

```
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2233 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  363..1778
US-08-496-631-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db    1055 TTGTCGCAAGTCAC 1042


RESULT  16
US-09-008-697A-7/c
; Sequence 7, Application US/09008697A
; Patent No. 6197504
;  GENERAL INFORMATION:
;    APPLICANT:  Chow, King Lau
;    TITLE OF INVENTION:  USES OF MAB-21
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,697A
;      FILING DATE:  January 19, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chan, Albert Wai-Kit
;      REGISTRATION NUMBER:  36,479
;      REFERENCE/DOCKET NUMBER:  50752-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2585 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-008-697A-7


  Query Match             56.0%;  Score 14;  DB 4;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatt 21
           ||||||||||||||
Db    1058 CAAGTCACAGAATT 1045
```

```
RESULT  17
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
US-08-765-081-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 AGTCACAGAATTAT 3740


RESULT  18
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
```

```
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;       COMPUTER:  IBM PC compatible/Pentium II
;       OPERATING SYSTEM:  MS-Windows 95
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/098,082
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/765,081
;       FILING DATE:  March 26, 1997
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sheiness, Diana K.
;       REGISTRATION NUMBER:  35,356
;       REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
US-09-098-082-1


  Query Match             56.0%;  Score 14;  DB 3;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 AGTCACAGAATTAT 3740


RESULT  19
PCT-US95-06994-1
; Sequence 1, Application PC/TUS9506994
;  GENERAL INFORMATION:
;     APPLICANT:  Children's Hospital & Medical Center
;     APPLICANT:  University of Washington
;     APPLICANT:  Washington State University Research Foundation
;     APPLICANT:  TARR, PHILLIP I
;     APPLICANT:  BILGE, SIMA S
;     APPLICANT:  BESSER, THOMAS E
;     APPLICANT:  VARY JR, JAMES C
;     TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;       STREET:  SUITE 2800, 1420 FIFTH AVENUE
;       CITY:  SEATTLE
;       STATE:  WASHINGTON
;       COUNTRY:  USA
;       ZIP:  WA 98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  07-JUN-95
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  24-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BRODERICK, THOMAS F
;       REGISTRATION NUMBER:  31,332
;       REFERENCE/DOCKET NUMBER:  CHOR-18591
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 682 8100
;       TELEFAX:  (206) 224 0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
PCT-US95-06994-1


  Query Match             56.0%;  Score 14;  DB 5;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 AGTCACAGAATTAT 3740


RESULT  20
US-08-875-847B-3/c
; Sequence 3, Application US/08875847B
; Patent No. 6255105
;  GENERAL INFORMATION:
;     APPLICANT:  The Government of the United
;     APPLICANT:  States of America as represented by the
;     APPLICANT:  Secretary, Department of Health and Human
;     APPLICANT:  Services; Callahan, Robert; Marchetti,
;     APPLICANT:  Antonio; Buttitta, Fiamma; Smith, Gilbert H.
;     TITLE OF INVENTION:  Nucleotide And Deduced
;     TITLE OF INVENTION:  Amino Acid Sequences Of A New Tumor Gene,
;     TITLE OF INVENTION:  Int6, And the Use Of Reagents Derived From
;     TITLE OF INVENTION:  These Sequences In Diagnostic Assays,
;     TITLE OF INVENTION:  Vaccines, Immunotherapy And Gene Therapy
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;       STREET:  345 PARK AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  MS WORD 97
;     CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/875,847B
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/385,998
;      FILING DATE:  09-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William S. Feiler
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4179PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-875-847B-3


  Query Match             52.0%;  Score 13;  DB 4;  Length 1500;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     624 TCACAGAATTATT 612


RESULT  21
US-08-875-847B-1/c
; Sequence 1, Application US/08875847B
; Patent No. 6255105
;  GENERAL INFORMATION:
;    APPLICANT:  The Government of the United
;    APPLICANT:  States of America as represented by the
;    APPLICANT:  Secretary, Department of Health and Human
;    APPLICANT:  Services; Callahan, Robert; Marchetti,
;    APPLICANT:  Antonio; Buttitta, Fiamma; Smith, Gilbert H.
;    TITLE OF INVENTION:  Nucleotide And Deduced
;    TITLE OF INVENTION:  Amino Acid Sequences Of A New Tumor Gene,
;    TITLE OF INVENTION:  Int6, And the Use Of Reagents Derived From
;    TITLE OF INVENTION:  These Sequences In Diagnostic Assays,
;    TITLE OF INVENTION:  Vaccines, Immunotherapy And Gene Therapy
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS WORD 97
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/875,847B
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/385,998
;      FILING DATE:  09-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William S. Feiler
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4179PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-875-847B-1


  Query Match             52.0%;  Score 13;  DB 4;  Length 1505;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     630 TCACAGAATTATT 618


RESULT  22
US-08-651-572-1/c
; Sequence 1, Application US/08651572
; Patent No. 5789228
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-651-572-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
            |||||||||||||
```

```
Db    974 AAGTCACAGAATT 962


RESULT  23
US-09-066-544-1/c
; Sequence 1, Application US/09066544
; Patent No. 6001984
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,544
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-066-544-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatt 21
          |||||||||||||
Db    974 AAGTCACAGAATT 962


RESULT  24
US-08-951-086-1/c
; Sequence 1, Application US/08951086
; Patent No. 6074867
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,086
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-951-086-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatt 21
          |||||||||||||
Db    974 AAGTCACAGAATT 962


RESULT  25
US-08-484-993B-1/c
; Sequence 1, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
```

```
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-484-993B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  26
US-08-484-158B-1/c
; Sequence 1, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Pharmaceutical Compositions for
;     TITLE OF INVENTION:  Immunocontraception
;     NUMBER OF SEQUENCES:  61
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;       ADDRESSEE:  Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/484,158B
;       FILING DATE:  07-JUNE-95
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/149,223
;       FILING DATE:  09-NOV-93
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-93
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-92
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  32794
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-484-158B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  27
US-08-484-596A-1/c
; Sequence 1, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/484,596A
```

```
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/149,223
;       FILING DATE:  11-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-484-596A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  28
US-08-480-150A-1/c
; Sequence 1, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-480-150A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  29
US-08-458-731-1/c
; Sequence 1, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
```

```
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,731
;       FILING DATE:  09-NOV-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-458-731-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  30
US-08-149-223A-1/c
; Sequence 1, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/149,223A
;       FILING DATE:  09-NOV-1993
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-149-223A-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  31
US-08-484-993B-9/c
; Sequence 9, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-993B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaa 19
          |||||||||||||
Db   2098 GCAAGTCACAGAA 2086


RESULT  32
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-158B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaa 19
          |||||||||||||
Db   2098 GCAAGTCACAGAA 2086


RESULT  33
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/484,596A
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/149,223
;       FILING DATE:  11-NOV-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2381 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Canis familiaris
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-484-596A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  34
US-08-480-150A-9/c
; Sequence 9, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/480,150A
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/149,223
;       FILING DATE:  09-NOV-1993
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2381 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Canis familiaris
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-480-150A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  35
US-08-458-731-9/c
; Sequence 9, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
```

```
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,731
;       FILING DATE:  09-NOV-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2381 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Canis familiaris
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-458-731-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  36
US-08-149-223A-9/c
; Sequence 9, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-149-223A-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  37
US-07-718-575-5/c
; Sequence 5, Application US/07718575
; Patent No. 5202257
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/718,575
;       FILING DATE:  19910813
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter Ph.D., Stephen E.
;       REGISTRATION NUMBER:  31192
;       REFERENCE/DOCKET NUMBER:  P31 8962
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619)535-8949
;       TELEX:  9103330318
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3083 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  GluR3
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  167..2833
US-07-718-575-5


  Query Match             52.0%;  Score 13;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtc 13
          |||||||||||||
Db   3033 ATTGTCGCAAGTC 3021


RESULT  38
US-08-481-206-5/c
; Sequence 5, Application US/08481206
; Patent No. 5739291
;  GENERAL INFORMATION:
;     APPLICANT:  Heinemann Ph.D., Stephen F.
;     APPLICANT:  Boulter Ph.D., James R.
;     APPLICANT:  Hollmann Ph.D., Michael NMN
;     APPLICANT:  Bettler Ph.D., Bernhard NMN
;     APPLICANT:  Jensen Ph.D., Jan E.
;     TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;     TITLE OF INVENTION:  METHODS
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;       STREET:  444 So. Flower St., Suite 2000
;       CITY:  Los Angeles
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  90071-2921
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/481,206
;       FILING DATE:
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/013,767
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter Ph.D., Stephen E.
;       REGISTRATION NUMBER:  31192
;       REFERENCE/DOCKET NUMBER:  P31 8962
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619)535-8949
;       TELEX:  9103330318
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3083 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  GluR3
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  167..2833
US-08-481-206-5


  Query Match             52.0%;  Score 13;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtc 13
          |||||||||||||
Db   3033 ATTGTCGCAAGTC 3021


RESULT  39
US-08-486-269A-5/c
; Sequence 5, Application US/08486269A
; Patent No. 5945509
;  GENERAL INFORMATION:
;     APPLICANT:  Heinemann, Stephen F.
;     APPLICANT:  Boulter, James R.
;     APPLICANT:  Hollmann, Michael
;     APPLICANT:  Bettler, Bernhard
;     APPLICANT:  Jensen, Jan E.
;     TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS
;     TITLE OF INVENTION:  AND METHODS
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;       STREET:  4365 Executive Drive, Suite 1600
;       CITY:  San Diego
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  92121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/486,269A
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/013,767
;       FILING DATE:  04-FEB-1993
;       APPLICATION NUMBER:  07/718,575
;       FILING DATE:  21-JUN-1991
;       APPLICATION NUMBER:  PCT/US90/06153
;       FILING DATE:  25-OCT-1990
;       APPLICATION NUMBER:  07/428,116
;       FILING DATE:  27-OCT-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter, Stephen E
;       REGISTRATION NUMBER:  31,192
;       REFERENCE/DOCKET NUMBER:  P41 9986
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  167...2830
;      OTHER INFORMATION:
US-08-486-269A-5


  Query Match             52.0%;  Score 13;  DB 2;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtc 13
          |||||||||||||
Db   3033 ATTGTCGCAAGTC 3021


RESULT  40
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
US-08-743-637B-35


  Query Match             52.0%;  Score 13;  DB 2;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatt 21
          |||||||||||||
Db   1911 AAGTCACAGAATT 1899


RESULT  41
US-08-526-840B-35/c
; Sequence 35, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,840B
;      FILING DATE:  11-SEP-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/304,732
;      FILING DATE:  12-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414) 277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
US-08-526-840B-35


  Query Match             52.0%;  Score 13;  DB 3;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaatt 21
```

```
          ||||||||||||||
Db   1911 AAGTCACAGAATT 1899


RESULT  42
US-09-050-159-129
; Sequence 129, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
;  APPLICANT: No. 6197505berg, Leif T
;  APPLICANT: Andersson, Maria K
;  APPLICANT: Linstrom, Per H
;  TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
;  TITLE OF INVENTION: COMPOSITIONS FOR USE THEREOF
;  FILE REFERENCE: 1248/1D042
;  CURRENT APPLICATION NUMBER: US/09/050,159A
;  CURRENT FILING DATE: 1998-03-27
;  EARLIER APPLICATION NUMBER: 60/042,930
;  EARLIER FILING DATE: 1987-04-03
;  NUMBER OF SEQ ID NOS: 133
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 129
;   LENGTH: 5590
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Angiotensin I converting enzyme, 5' region
US-09-050-159-129


  Query Match             52.0%;  Score 13;  DB 4;  Length 5590;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     294 tcacagaattatt 306


RESULT  43
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  72928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db   26365 AAGTCACAGAATT 26353


RESULT  44
US-08-373-124A-284
; Sequence 284, Application US/08373124A
; Patent No. 5646042
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  Draper, Kenneth
;    APPLICANT:  McSwiggen, James
;    APPLICANT:  Jarvis, Thale
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR
;    TITLE OF INVENTION:  TREATMENT OF RESTENOSIS AND
;    TITLE OF INVENTION:  CANCER USING RIBOZYMES
;    NUMBER OF SEQUENCES:  2627
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/373,124A
;      FILING DATE:  January 13, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/245,466
;      FILING DATE:  May 18, 1994
;      APPLICATION NUMBER:  08/192,943
;      FILING DATE:  February 7, 1994
;      APPLICATION NUMBER:  07/987,132
;      FILING DATE:  December 7, 1992
;      APPLICATION NUMBER:  07/936,422
;      FILING DATE:  August 26, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  209/035
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  284:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

US-08-373-124A-284

Query Match 48.0%; Score 12; DB 1; Length 17;
Best Local Similarity 58.3%; Pred. No. 2e+02;
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy      14 acagaattattt 25
           ||||||::|:::
Db       2 ACAGAAUUAUUU 13
```

RESULT 45
US-08-373-124A-286
; Sequence 286, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 286:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-373-124A-286

Query Match 48.0%; Score 12; DB 1; Length 17;
Best Local Similarity 58.3%; Pred. No. 2e+02;
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```
Qy      14 acagaattattt 25
           ||||||::|:::
Db       1 ACAGAAUUAUUU 12
```

Search completed: October 9, 2001, 15:55:21
Job time: 13197 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:21 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 563573

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	18	72.0	357	140	BE800706	BE800706 sq97c06.y
2	18	72.0	388	120	AW781150	AW781150 sl89g06.y
c 3	18	72.0	751	138	BE658717	BE658717 GM700007A
4	17	68.0	424	225	AQ227679	AQ227679 HS_2012_B
c 5	17	68.0	944	21	AI529608	AI529608 ui80b10.y
c 6	16	64.0	263	126	BB141316	BB141316 BB141316
c 7	16	64.0	270	164	BE181705	BE181705 QV1-HT063
8	16	64.0	299	134	BB450863	BB450863 BB450863
9	16	64.0	353	10	AA648449	AA648449 ns22f10.s
10	16	64.0	405	168	BF726110	BF726110 by01c01.x
11	16	64.0	429	146	BF232119	BF232119 dd99d06.x
12	16	64.0	443	16	AI149091	AI149091 qc85f02.x
13	16	64.0	454	224	AQ086771	AQ086771 HS_2175_B
14	16	64.0	457	19	AI361386	AI361386 qy47f01.x
15	16	64.0	467	152	BG360303	BG360303 dab93h11.
16	16	64.0	471	22	AI635259	AI635259 tz79h07.x
17	16	64.0	503	22	AI634722	AI634722 tz09e04.x
18	16	64.0	503	152	BG347513	BG347513 dab87h12.
19	16	64.0	512	115	AW418528	AW418528 xr32c07.x
20	16	64.0	527	229	AQ472309	AQ472309 CITBI-E1-
c 21	16	64.0	649	147	BF322232	BF322232 maa34h05.
c 22	16	64.0	652	230	AQ541562	AQ541562 RPCI-11-3
c 23	16	64.0	657	168	BF726111	BF726111 by01c01.y
24	16	64.0	702	242	AZ382380	AZ382380 1M0139H04
c 25	16	64.0	753	154	BG496183	BG496183 602538125
c 26	16	64.0	971	235	AQ900023	AQ900023 HS_3180_A
c 27	15	60.0	214	121	AW871559	AW871559 ra79d04.y
28	15	60.0	233	139	BE763893	BE763893 RC4-NT005
29	15	60.0	253	139	BE763892	BE763892 RC4-NT005
30	15	60.0	256	5	AA317110	AA317110 EST19001
c 31	15	60.0	261	191	Z45844	Z45844 HSCZVE041 n
32	15	60.0	263	169	BF745516	BF745516 CM2-BT082
c 33	15	60.0	266	2	AA115083	AA115083 zl06c10.r
c 34	15	60.0	282	169	BF800965	BF800965 QV4-CI010
35	15	60.0	286	6	AA344205	AA344205 EST50086
c 36	15	60.0	292	256	B53227	B53227 CIT-HSP-200
c 37	15	60.0	317	169	BF743951	BF743951 CM2-BT082
c 38	15	60.0	327	188	T30901	T30901 EST24536 Hu
c 39	15	60.0	335	169	BF743964	BF743964 CM2-BT082
c 40	15	60.0	346	169	BF743954	BF743954 CM2-BT082
41	15	60.0	370	188	T07051	T07051 EST04940 Fe
c 42	15	60.0	381	158	H53286	H53286 yq85f09.r1
c 43	15	60.0	395	4	AA229851	AA229851 nc51e03.r
44	15	60.0	402	23	AI694429	AI694429 wd83f06.x
45	15	60.0	404	17	AI203847	AI203847 qf75h03.x

ALIGNMENTS

```
RESULT   1
BE800706
LOCUS       BE800706      357 bp    mRNA            EST       20-SEP-2000
DEFINITION  sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1049-1091 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH
            PROTEASE. ;, mRNA sequence.
ACCESSION   BE800706
VERSION     BE800706.1  GI:10231818
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Putative full length read
            vector to vector length is 364
            High quality sequence stop: 340.
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
                     /clone_lib="Gm-c1049"
                     /tissue_type="whole seedlings of greenhouse grown plants"
                     /dev_stage="3 week old"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Clark NIL was constructed and seed was provided
                     by Dr. J. Specht, University of Nebraska (Shoemaker and
                     Specht, 1995). The cDNA library was constructed from mRNA
                     isolated from whole seedlings of 3 week old greenhouse
                     grown plants. Complementary DNA was synthesized from mRNA
                     using a primer consisting of a poly(dT) sequence with a
                     XhoI restriction site and a 3' anchor. EcoRI adapters were
                     ligated to the blunt-ended cDNA fragments followed by XhoI
                     digestion. The cDNA fragments were directionally cloned
                     into the EcoRI-XhoI restriction site of the pBluescript
                     vector. The ligated cDNA fragments were transformed into
                     DH10B host cells (GibcoBRL). The library was constructed
                     in cooperation with Dr. Paul Keim's laboratory at Northern
                     Arizona University."
BASE COUNT      169 a     53 c     47 g     88 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 140;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 cgcaagtcacagaattat 23
           ||||||||||||||||||
Db     138 CGCAAGTCACAGAATTAT 155


RESULT   2
AW781150
LOCUS       AW781150      388 bp    mRNA            EST       12-MAY-2000
DEFINITION  sl89g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1037-1235 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH
            PROTEASE. ;, mRNA sequence.
ACCESSION   AW781150
VERSION     AW781150.1  GI:7795753
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 388)
```

```
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bqlla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1037-1235"
                     /clone_lib="Gm-c1037"
                     /tissue_type="fully expanded leaves of greenhouse grown
                     plants"
                     /dev_stage="2 week old"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from fully
                     expanded leaves of greenhouse grown plants that were 2
                     weeks old. The library was prepared using the Life
                     Technologies pSuperScript cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a
                     poly(dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the  blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli Electro- Max DH10B host
                     cells. This library was constructed in the laboratory of
                     Dr. Lila Vodkin by Anu Khanna at the University of
                     Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"
BASE COUNT      123 a     67 c     92 g    106 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 120;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 cgcaagtcacagaattat 23
          ||||||||||||||||||
Db    334 CGCAAGTCACAGAATTAT 351


RESULT   3
BE658717/c
LOCUS       BE658717      751 bp    mRNA            EST       06-SEP-2000
DEFINITION  GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',
            mRNA sequence.
ACCESSION   BE658717
VERSION     BE658717.1  GI:9984609
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW156684 corresponding to Gm-c1015-2632 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-2561"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID of
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
BASE COUNT      193 a    171 c    131 g    226 t     30 others
ORIGIN


  Query Match             72.0%;  Score 18;  DB 138;  Length 751;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 cgcaagtcacagaattat 23
          ||||||||||||||||||
Db    112 CGCAAGTCACAGAATTAT 95


RESULT   4
AQ227679
LOCUS       AQ227679      424 bp    DNA             GSS       26-SEP-1998
DEFINITION  HS_2012_B2_F10_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2012 Col=20 Row=L, DNA sequence.
ACCESSION   AQ227679
VERSION     AQ227679.1  GI:3652908
KEYWORDS    GSS.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2012  row: L  column: 20
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2012 Col=20 Row=L"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      107 a     86 c     87 g    143 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 225;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattattt 25
          |||||||||||||||||
Db    250 AAGTCACAGAATTATTT 266


RESULT  5
AI529608/c
LOCUS       AI529608      944 bp    mRNA            EST       18-MAR-1999
DEFINITION  ui80b10.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1888699 5' similar to gb:L11650 Mouse cytochrome P-450 IIE1
            mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION   AI529608
VERSION     AI529608.1  GI:4443743
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:973023
            Seq primer: custom primer used
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1888699"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      253 a    238 c    226 g    227 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 21;  Length 944;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattattt 25
          |||||||||||||||||
Db    906 AAGTCACAGAATTATTT 890


RESULT  6
BB141316/c
LOCUS       BB141316      263 bp    mRNA            EST       28-JUN-2000
DEFINITION  BB141316 RIKEN full-length enriched, adult female vagina Mus
            musculus cDNA clone 9930007O04 3', mRNA sequence.
ACCESSION   BB141316
VERSION     BB141316.1  GI:8796253
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
```

Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9930007004"
                     /clone_lib="RIKEN full-length enriched, adult female
                     vagina"
                     /sex="female"
                     /tissue_type="vagina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT       69 a     56 c     43 g     95 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 126;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db      62 CAAGTCACAGAATTAT 47
```

```
RESULT   7
BE181705/c
LOCUS       BE181705      270 bp    mRNA            EST       22-JUN-2000
DEFINITION  QV1-HT0638-080500-197-d06 HT0638 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE181705
VERSION     BE181705.1  GI:8660881
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 270)
  AUTHORS   Dias Neto,E., Garcia-Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-HT0638-080
            500-197-d06&t3=2000-05-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 111
            High quality sequence stop: 223.
FEATURES             Location/Qualifiers
     source          1. .270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0638"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       89 a     57 c     53 g     71 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 164;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db     227 AAGTCACAGAATTATT 212
```

```
RESULT   8
BB450863
LOCUS       BB450863      299 bp    mRNA            EST       21-JUL-2000
DEFINITION  BB450863 RIKEN full-length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130006I17 3', mRNA sequence.
ACCESSION   BB450863
VERSION     BB450863.1  GI:9311898
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
```

```
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D130006I17"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal ganglion"
                     /tissue_type="spinal ganglion"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       75 a     72 c     45 g    107 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 134;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 cgcaagtcacagaatt 21
           ||||||||||||||||
Db       3 CGCAAGTCACAGAATT 18


RESULT   9
AA648449
LOCUS       AA648449      353 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns22f10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
            similar to contains element LTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AA648449
VERSION     AA648449.1  GI:2574878
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Insert Length: 2027   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 334.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1184395"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      131 a     49 c     39 g    134 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 353;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     249 CAAGTCACAGAATTAT 264


RESULT  10
BF726110
LOCUS       BF726110      405 bp    mRNA            EST       05-JAN-2001
DEFINITION  by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION   BF726110
VERSION     BF726110.1  GI:12042021
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
```

```
            Email: graeme@helix.nih.gov
            Plate: 01  row: c  column: 01
            Seq primer: -21M13 forward primer (ABI).
FEATURES              Location/Qualifiers
     source           1. .405
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="by01c01"
                      /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                      BY"
                      /tissue_type="Lens"
                      /dev_stage="Adult"
                      /lab_host="EMDH10B"
                      /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                      from different adults (both approximately 40 years old)
                      together yielded 20ug of total RNA and 150ng mRNA for cDNA
                      library synthesis. A directionally cloned cDNA library in
                      the pCMVSPORT6 vector was constructed at Life Technologies
                      , essentially following the protocols of the SuperScript
                      Plasmid System full details of which are contained in the
                      manufacturer's Instruction manual
                      (http://www.lifetech.com/). First strand synthesis was
                      carried out using a Not I primer-adapter
                      [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                      end inserts were cloned into the Not I/EcoR V sites in the
                      vector. EST analysis was performed on the unamplified
                      library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT      140 a     58 c     47 g    160 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 168;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     249 CAAGTCACAGAATTAT 264


RESULT  11
BF232119
LOCUS       BF232119      429 bp    mRNA            EST       11-DEC-2000
DEFINITION  dd99d06.x1 Wellcome CRC pRN3 oocyte Xenopus laevis cDNA clone
            IMAGE:3436475 3', mRNA sequence.
ACCESSION   BF232119
VERSION     BF232119.1  GI:11174440
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
            University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            High quality sequence stop: 425.
FEATURES              Location/Qualifiers
     source           1. .429
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:3436475"
                      /clone_lib="Wellcome CRC pRN3 oocyte"
                      /tissue_type="oocyte"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                      were oligo-dT primed and directionally cloned. Staging
                      according to  Nieuwkoop and Faber. Library was constructed
                      by N. Garrett, P. LeMaire, A.M. Zorn,  and J.B. Gurdon
                      (Wellcome/CRC Institute). "
BASE COUNT      141 a     73 c     74 g    141 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 146;  Length 429;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db      60 AGTCACAGAATTATTT 75


RESULT  12
AI149091
LOCUS       AI149091      443 bp    mRNA            EST       28-OCT-1998
DEFINITION  qc85f02.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1721019 3', mRNA sequence.
ACCESSION   AI149091
VERSION     AI149091.1  GI:3677560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 527   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .443
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1721019"
                      /clone_lib="Soares_pregnant_uterus_NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
BASE COUNT      126 a    122 c     71 g    124 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 16;  Length 443;
  Best Local Similarity   100.0%;  Pred. No. 26;
```

```
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db     100 AGTCACAGAATTATTT 115


RESULT  13
AQ086771
LOCUS       AQ086771      454 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2175_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION   AQ086771
VERSION     AQ086771.1  GI:3455988
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2175  row: B  column: 12
            Class: BAC ends
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2175 Col=12 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      188 a     70 c     76 g    117 t      3 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 224;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    156 CAAGTCACAGAATTAT 171


RESULT  14
AI361386
LOCUS       AI361386      457 bp    mRNA            EST       15-FEB-1999
DEFINITION  qy47f01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2015161 3',
            mRNA sequence.
ACCESSION   AI361386
VERSION     AI361386.1  GI:4113007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1228   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2015161"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      153 a     73 c     52 g    179 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 19;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    153 CAAGTCACAGAATTAT 168


RESULT  15
BG360303
LOCUS       BG360303      467 bp    mRNA            EST       06-MAR-2001
DEFINITION  dab93h11.y1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4405125
            5', mRNA sequence.
ACCESSION   BG360303
VERSION     BG360303.1  GI:13240994
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
```

```
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Library constructed by Life Technologies. DNA Sequencing by:
           Washington University Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
           through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
           Seq primer: -40RP from Gibco
           High quality sequence stop: 465.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4405125"
                     /clone_lib="NICHD XGC Emb2"
                     /tissue_type="embryo, stage 17/19"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by Life Technologies."
BASE COUNT      146 a     86 c     84 g    151 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 152;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db      89 AGTCACAGAATTATTT 104


RESULT  16
AI635259
LOCUS       AI635259      471 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294845 3',
            mRNA sequence.
ACCESSION   AI635259
VERSION     AI635259.1  GI:4686589
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1533   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2294845"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      171 a     69 c     53 g    178 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 22;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     235 CAAGTCACAGAATTAT 250


RESULT  17
AI634722
LOCUS       AI634722      503 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz09e04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2288094 3'
            similar to contains element XTR repetitive element ;, mRNA
            sequence.
ACCESSION   AI634722
VERSION     AI634722.1  GI:4686052
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3719   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2288094"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      184 a     77 c     56 g    186 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 22;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     236 CAAGTCACAGAATTAT 251


RESULT  18
BG347513
LOCUS       BG347513      503 bp    mRNA            EST       28-FEB-2001
DEFINITION  dab87h12.x1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4404719
```

```
            3', mRNA sequence.
ACCESSION   BG347513
VERSION     BG347513.1  GI:13167937
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4404719"
                     /clone_lib="NICHD XGC Emb2"
                     /tissue_type="embryo, stage 17/19"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by Life Technologies."
BASE COUNT      167 a     87 c     89 g    160 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 152;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db      51 AGTCACAGAATTATTT 66


RESULT  19
AW418528
LOCUS       AW418528      512 bp    mRNA            EST       09-FEB-2000
DEFINITION  xr32c07.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2761836 3'
            similar to contains element MER32 repetitive element ;, mRNA
            sequence.
ACCESSION   AW418528
VERSION     AW418528.1  GI:6946410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2761836"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT      186 a     79 c     56 g    191 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 115;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     240 CAAGTCACAGAATTAT 255


RESULT  20
AQ472309
LOCUS       AQ472309      527 bp    DNA             GSS       23-APR-1999
DEFINITION  CITBI-E1-2587F18.TF CITBI-E1 Homo sapiens genomic clone 2587F18,
            DNA sequence.
ACCESSION   AQ472309
VERSION     AQ472309.1  GI:4655963
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2587F18.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2587F18"
                     /clone_lib="CITBI-E1"
```

```
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT       150 a     88 c     83 g    205 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 229;  Length 527;
  Best Local Similarity   100.0%;  Pred. No. 26;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db      81 AGTCACAGAATTATTT 96


RESULT  21
BF322232/c
LOCUS       BF322232      649 bp    mRNA            EST       21-NOV-2000
DEFINITION  maa34h05.y1 NCI_CGAP_Li10 Mus musculus cDNA clone IMAGE:3812985 5'
            similar to SW:CPE1_MOUSE Q05421 CYTOCHROME P450 2E1 ;, mRNA
            sequence.
ACCESSION   BF322232
VERSION     BF322232.1  GI:11271563
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            MGI:1455097
            Seq primer: -40RP from Gibco
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3812985"
                     /clone_lib="NCI_CGAP_Li10"
                     /sex="female"
                     /dev_stage="10 weeks"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.6 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       201 a    150 c    150 g    147 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 147;  Length 649;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db     517 AGTCACAGAATTATTT 502


RESULT  22
AQ541562/c
LOCUS       AQ541562      652 bp    DNA             GSS       19-MAY-1999
DEFINITION  RPCI-11-361M22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-361M22
            , DNA sequence.
ACCESSION   AQ541562
VERSION     AQ541562.1  GI:4872018
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-361M22.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Homo sapiens"
                     /db_xref="GDB:7638549"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-361M22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT       201 a    162 c     82 g    207 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 230;  Length 652;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db      96 CAAGTCACAGAATTAT 81


RESULT  23
BF726111/c
LOCUS       BF726111      657 bp    mRNA            EST       05-JAN-2001
DEFINITION  by01c01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by01c01 5', mRNA sequence.
ACCESSION   BF726111
VERSION     BF726111.1  GI:12042022
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
```

```
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 01  row: c  column: 01
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="by01c01"
                     /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                     BY"
                     /tissue_type="Lens"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                     from different adults (both approximately 40 years old)
                     together yielded 20ug of total RNA and 150ng mRNA for cDNA
                     library synthesis. A directionally cloned cDNA library in
                     the pCMVSPORT6 vector was constructed at Life Technologies
                     , essentially following the protocols of the SuperScript
                     Plasmid System full details of which are contained in the
                     manufacturer's Instruction manual
                     (http://www.lifetech.com/). First strand synthesis was
                     carried out using a Not I primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                     end inserts were cloned into the Not I/EcoR V sites in the
                     vector. EST analysis was performed on the unamplified
                     library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT      232 a     86 c    114 g    225 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 168;  Length 657;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    574 CAAGTCACAGAATTAT 559


RESULT  24
AZ382380
LOCUS       AZ382380      702 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0139H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0139H04 R, DNA sequence.
ACCESSION   AZ382380
VERSION     AZ382380.1  GI:10496080
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0139  row: H  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 702.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0139H04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      272 a    120 c    104 g    206 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 242;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattatt 24
          ||||||||||||||||
Db    637 AAGTCACAGAATTATT 652


RESULT  25
BG496183/c
LOCUS       BG496183      753 bp    mRNA            EST       27-MAR-2001
DEFINITION  602538125F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4658982 5',
            mRNA sequence.
ACCESSION   BG496183
VERSION     BG496183.1  GI:13457699
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
            http://image.llnl.gov
            Plate: LLCM1454  row: b  column: 07
            High quality sequence start: 20
            High quality sequence stop: 694.
FEATURES              Location/Qualifiers
     source           1. .753
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4658982"
                      /clone_lib="NIH_MGC_59"
                      /tissue_type="mucoepidermoid carcinoma"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                      SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                      Double-stranded cDNA was prepared from cell line RNA.  5'
                      and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C,  G, or T).  Average
                      insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                      contained  inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                      Library."
BASE COUNT      202 a    143 c    174 g    233 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 154;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattatt 24
          ||||||||||||||||
Db     46 AAGTCACAGAATTATT 31


RESULT  26
AQ900023/c
LOCUS       AQ900023      971 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_3180_A2_G11_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3180 Col=22 Row=M, DNA sequence.
ACCESSION   AQ900023
VERSION     AQ900023.1  GI:6356213
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3180  row: M  column: 22
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 971.
FEATURES              Location/Qualifiers
     source           1. .971
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=3180 Col=22 Row=M"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      249 a    130 c    128 g    464 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 235;  Length 971;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 agtcacagaattattt 25
          ||||||||||||||||
Db    501 AGTCACAGAATTATTT 486


RESULT  27
AW871559/c
LOCUS       AW871559      214 bp    mRNA            EST       05-APR-2001
DEFINITION  ra79d04.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
            cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION   AW871559
VERSION     AW871559.1  GI:8005612
KEYWORDS    EST.
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 214)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Uma Rao and David Bird
            (david_bird@ncsu.edu) at North Carolina State University. DNA
            Sequencing by: Washington University Genome Sequencing Center St.
            Louis.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .214
                      /organism="Meloidogyne incognita"
                      /db_xref="taxon:6306"
                      /clone_lib="Bird-Rao Meloidogyne incognita J2"
                      /dev_stage="enriched for 2nd stage juveniles"
                      /lab_host="XLORL"
                      /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:
                      EcoRI; Site_2: XhoI; Oligo (dT) primed library.  cDNA was
                      contructed and cloned  unidirectionally into the vector
                      within the 5' EcoRI and  3' XhoI sites.  This library was
                      constructed by Dr. Uma Rao  and Dr. David Bird at North
                      Carolina state University."
BASE COUNT       59 a     52 c     19 g     84 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 121;  Length 214;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     184 CAAGTCACAGAATTA 170


RESULT  28
BE763893
LOCUS       BE763893      233 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763893
VERSION     BE763893.1  GI:10193817
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-NT0054-120
            600-015-g05&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 233.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a     27 c     52 g     66 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 139;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db      94 CAAGTCACAGAATTA 108


RESULT  29
BE763892
LOCUS       BE763892      253 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763892
VERSION     BE763892.1  GI:10193816
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-NT0054-120
            600-015-g03&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       95 a     31 c     56 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 139;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     114 CAAGTCACAGAATTA 128


RESULT  30
AA317110
LOCUS       AA317110      256 bp    mRNA            EST       19-APR-1997
DEFINITION  EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA317110
VERSION     AA317110.1  GI:1969448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256)
```

```
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .256
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):117571"
                      /db_xref="taxon:9606"
                      /clone_lib="Lung"
                      /dev_stage="adult"
                      /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                      ; Site_2: XhoI"
BASE COUNT       73 a     57 c     48 g     75 t      3 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 5;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     189 CAAGTCACAGAATTA 203


RESULT  31
Z45844/c
LOCUS       Z45844        261 bp    mRNA            EST       14-NOV-1994
DEFINITION  HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone
            c-zve04, mRNA sequence.
ACCESSION   Z45844
VERSION     Z45844.1  GI:575078
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-zve04
            Seq primer: (-21)M13_universal.
FEATURES              Location/Qualifiers
     source           1. .261
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="c-zve04"
                      /clone_lib="normalized infant brain cDNA"
                      /sex="Female"
                      /tissue_type="total brain"
                      /dev_stage="3 months old"
                      /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                      Site_2: NotI; sex=Female; dev_stage=3 months old;
                      isolate=muscular atrophy patient;  tissue_type=total brain
                      ;  total mRNA was oligo-(dT) primed and directionally
                      cloned 5' -> 3' into the  HindIII -> NotI sites of the
                      lafmid BA vector. Clone library from B.Soares,  Psychiatry
                      Dept. Columbia University, USA.  Normalization_method:
                      Bento Soares, P.N.A.S in press"
BASE COUNT       77 a     56 c     31 g     97 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 191;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     138 CAAGTCACAGAATTA 124


RESULT  32
BF745516
LOCUS       BF745516      263 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-021100-469-h05 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF745516
VERSION     BF745516.1  GI:12072192
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
```

```
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            021100-469-h05&t3=2000-11-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 129
            High quality sequence stop: 263.
FEATURES              Location/Qualifiers
     source           1. .263
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT       77 a     55 c     60 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db    217 GCAAGTCACAGAATT 231


RESULT  33
AA115083/c
LOCUS       AA115083      266 bp    mRNA            EST       14-MAY-1997
DEFINITION  zl06c10.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:491538 5', mRNA sequence.
ACCESSION   AA115083
VERSION     AA115083.1  GI:1670318
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 555   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES              Location/Qualifiers
     source           1. .266
                      /organism="Homo sapiens"
                      /db_xref="GDB:3806125"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:491538"
                      /clone_lib="Soares_pregnant_uterus_NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
BASE COUNT       61 a     67 c     64 g     69 t      5 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 2;  Length 266;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db      91 GTCACAGAATTATTT 77


RESULT  34
BF800965/c
LOCUS       BF800965      282 bp    mRNA            EST       12-JAN-2001
DEFINITION  QV4-CI0100-231000-457-d12 CI0100 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF800965
VERSION     BF800965.1  GI:12129954
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV4&t2=QV4-CI0100-
            231000-457-d12&t3=2000-10-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 282.
FEATURES              Location/Qualifiers
     source           1. .282
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="CI0100"
                      /dev_stage="Adult"
                      /note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
```

```
                    low stringency conditions."
BASE COUNT       68 a     75 c     67 g     72 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 282;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcac 15
          |||||||||||||||
Db     51 ATTGTCGCAAGTCAC 37


RESULT  35
AA344205
LOCUS       AA344205       286 bp    mRNA            EST       21-APR-1997
DEFINITION  EST50086 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA344205
VERSION     AA344205.1  GI:1996443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):145738"
                     /db_xref="taxon:9606"
                     /clone_lib="Gall bladder I"
                     /sex="female"
                     /dev_stage="adult, 25 yrs"
                     /note="Organ: gall bladder; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       83 a     62 c     48 g     91 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          |||||||||||||||
Db    197 CAAGTCACAGAATTA 211


RESULT  36
B53227/c
LOCUS       B53227         292 bp    DNA             GSS       20-JUN-1998
DEFINITION  CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
            sequence.
ACCESSION   B53227
VERSION     B53227.1  GI:2607561
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="GDB:7040812"
                     /db_xref="taxon:9606"
                     /clone="2009H2"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       58 a     64 c     90 g     80 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 256;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          |||||||||||||||
Db    248 CAAGTCACAGAATTA 234


RESULT  37
BF743951/c
LOCUS       BF743951       317 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743951
VERSION     BF743951.1  GI:12070627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-c02&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       96 a     71 c     49 g    101 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     45 GCAAGTCACAGAATT 31


RESULT  38
T30901/c
LOCUS       T30901         327 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
            sequence.
ACCESSION   T30901
VERSION     T30901.1  GI:612999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
            Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
            Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC10632
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):100882"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Brain"
                     /note="Organ: brain"
BASE COUNT       89 a     75 c     47 g    115 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 188;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          |||||||||||||||
Db    295 CAAGTCACAGAATTA 281


RESULT  39
BF743964/c
LOCUS       BF743964       335 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743964
VERSION     BF743964.1  GI:12070640
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
```

```
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-h04&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 40
            High quality sequence stop: 335.
FEATURES             Location/Qualifiers
     source          1. .335
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      101 a     76 c     55 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 335;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     61 GCAAGTCACAGAATT 47


RESULT  40
BF743954/c
LOCUS       BF743954      346 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743954
VERSION     BF743954.1  GI:12070630
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-d07&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 43
            High quality sequence stop: 346.
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      103 a     78 c     54 g    111 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 169;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     73 GCAAGTCACAGAATT 59


RESULT  41
T07051
LOCUS       T07051        370 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST04940 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBEC79, mRNA sequence.
ACCESSION   T07051
VERSION     T07051.1  GI:318200
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):83698"
                     /db_xref="taxon:9606"
                     /clone="HFBEC79"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT      137 a     58 c     68 g    106 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 188;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          |||||||||||||||
Db    113 CAAGTCACAGAATTA 127
```

```
RESULT  42
H53286/c
LOCUS       H53286         381 bp    mRNA            EST       20-SEP-1995
DEFINITION  yq85f09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:202601 5', mRNA sequence.
ACCESSION   H53286
VERSION     H53286.1  GI:993433
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1459
            High quality sequence stops: 288
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1459   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="Homo sapiens"
                     /db_xref="GDB:3771732"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:202601"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      105 a     99 c     79 g     88 t     10 others
ORIGIN


  Query Match           60.0%;  Score 15;  DB 158;  Length 381;
  Best Local Similarity  100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db     281 GTCACAGAATTATTT 267


RESULT  43
AA229851/c
LOCUS       AA229851       395 bp    mRNA            EST       21-AUG-1997
DEFINITION  nc51e03.r1 NCI_CGAP_Pr3 Homo sapiens cDNA clone IMAGE:1011676, mRNA
            sequence.
ACCESSION   AA229851
VERSION     AA229851.1  GI:1852164
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
            , Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 622   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1011676"
                     /clone_lib="NCI_CGAP_Pr3"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: Not1; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from 5,000-10
                     ,000 microdissected cells histologically-determined to be
                     fully malignant prostate cancer cells.  Double-stranded
                     cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
                     applied to the cDNA with an adaptor-specific primer, and
                     the resulting PCR product subcloned into pAMP10 by the
                     UDG-cloning method (Life Technologies).  Average insert
                     size is 600 bp. NOTE: Not directionally cloned.  This
                     library was constructed by  David Krizman."
BASE COUNT      110 a     62 c     65 g     158 t
ORIGIN


  Query Match           60.0%;  Score 15;  DB 4;  Length 395;
  Best Local Similarity  100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattat 23
           |||||||||||||||
Db      83 AAGTCACAGAATTAT 69


RESULT  44
AI694429
LOCUS       AI694429       402 bp    mRNA            EST       02-JUN-1999
DEFINITION  wd83f06.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338211 3',
            mRNA sequence.
ACCESSION   AI694429
VERSION     AI694429.1  GI:4971769
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2338211"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      160 a     60 c     69 g    113 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 23;  Length 402;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     342 CAAGTCACAGAATTA 356


RESULT  45
AI203847
LOCUS       AI203847      404 bp    mRNA            EST       27-JAN-1999
DEFINITION  qf75h03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755893
            3', mRNA sequence.
ACCESSION   AI203847
VERSION     AI203847.1  GI:3756453
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 563   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1755893"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      161 a     61 c     69 g    113 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 17;  Length 404;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     342 CAAGTCACAGAATTA 356


Search completed: October  9, 2001, 15:15:23
Job time: 13645 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:44:37 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 254803

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	128	2	ECOBIOB	M27731 Escherichia
2	25	100.0	1041	9	AR029499	AR029499 Sequence
3	25	100.0	1041	9	AR034916	AR034916 Sequence
4	25	100.0	1084	9	A11530	A11530 BioB gene o
5	25	100.0	5526	2	AF250776	AF250776 Unculture
6	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
7	25	100.0	5872	9	A38246	A38246 Sequence 1
8	25	100.0	5872	9	A38251	A38251 Sequence 6

	9	25	100.0	5872	9	A93674 A93674 Sequence 1
	10	25	100.0	5872	9	A93679 A93679 Sequence 6
	11	25	100.0	5872	9	AR101809 AR101809 Sequence
	12	25	100.0	5872	9	AR101810 AR101810 Sequence
	13	25	100.0	11022	1	AE000180 AE000180 Escherich
	14	25	100.0	13501	1	AE005258 AE005258 Escherich
	15	25	100.0	297816	2	AP002553 AP002553 Escherich
	16	24	96.0	1121	10	E00893 E00893 Genomic DNA
	17	20	80.0	12891	1	AE004192 AE004192 Vibrio ch
c	18	17	68.0	383	96	S64534 S64534 Drosophila
c	19	17	68.0	461	96	S64532 S64532 Drosophila
c	20	17	68.0	516	6	DMDECG1B X76231 D.melanogas
c	21	17	68.0	539	96	S64530 S64530 Drosophila
c	22	17	68.0	539	96	S64531 S64531 Drosophila
c	23	17	68.0	617	96	S64529 S64529 Drosophila
c	24	17	68.0	774	6	DSDEC1TR6 X96931 D.sechellia
c	25	17	68.0	852	6	DSDEC1TR2 X96927 D.simulans
c	26	17	68.0	852	6	DSDEC1TR3 X96928 D.simulans
c	27	17	68.0	930	6	DMDEC1TR1 X96926 D.melanogas
c	28	17	68.0	3741	6	DRODEC1A M35887 Drosophila
c	29	17	68.0	3755	6	DRODEC1B M35888 Drosophila
c	30	17	68.0	5697	6	DRODEC1C M35889 Drosophila
c	31	17	68.0	33204	65	AC017995 AC017995 Drosophil
c	32	17	68.0	161540	64	AC016952 AC016952 Homo sapi
	33	17	68.0	306267	4	AE003442 AE003442 Drosophil
c	34	16	64.0	774	6	DMDEC1TR5 X96930 D.mauritian
c	35	16	64.0	1614	8	AF281080 AF281080 Xenopus l
	36	16	64.0	5000	13	AF143940 AF143940 Arabidops
	37	16	64.0	6887	15	YLI012084 AJ012084 Yarrowia
c	38	16	64.0	11328	1	AE004322 AE004322 Vibrio ch
	39	16	64.0	13856	1	AE004026 AE004026 Xylella f
	40	16	64.0	15247	1	AE004580 AE004580 Pseudomon
	41	16	64.0	31557	92	HS314G4 Z69667 Human DNA s
	42	16	64.0	41200	14	SC8520X Z49705 S.cerevisia
c	43	16	64.0	46340	9	AX008043 AX008043 Sequence
	44	16	64.0	62927	12	AB018119 AB018119 Arabidops
c	45	16	64.0	65679	63	AC013956 AC013956 Drosophil

ALIGNMENTS

```
RESULT   1
ECOBIOB
LOCUS       ECOBIOB       128 bp    DNA             BCT       03-JAN-1995
DEFINITION  Escherichia coli biotin (bioB) gene, early terminator region.
ACCESSION   M27731
VERSION     M27731.1  GI:341755
KEYWORDS    bioB gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Nath,S.K.
  TITLE     Attenuation of transcription of biotin genes in Escherichia coli
  JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
  MEDLINE   89167942
FEATURES             Location/Qualifiers
     source          1. .128
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     gene            4. .128
                     /gene="bioB"
     -10_signal      4. .9
                     /gene="bioB"
     mRNA            20. .>128
                     /gene="bioB"
     terminator      103. .121
                     /gene="bioB"
                     /note="early terminator"
BASE COUNT       31 a     28 c     34 g     35 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 2;  Length 128;
  Best Local Similarity  100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      18 GTTGCTGGATCTGCTGTTTGAAGCG 42


RESULT   2
AR029499
LOCUS       AR029499     1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   3
AR034916
LOCUS       AR034916     1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   4
```

```
A11530
LOCUS       A11530       1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     80 GTTGCTGGATCTGCTGTTTGAAGCG 104


RESULT   5
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
```

```
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a    1507 c    1567 g    1178 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 2;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches    25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db    2019 GTTGCTGGATCTGCTGTTTGAAGCG 2043


RESULT   6
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
```

```
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 2;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   2068 GTTGCTGGATCTGCTGTTTGAAGCG 2092


RESULT   7
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT   8
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT   9
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
```

```
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  10
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  11
AR101809
```

```
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  12
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 9;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  13
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
     CDS             complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
```

```
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQQTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   3549 GTTGCTGGATCTGCTGTTTGAAGCG 3573


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
```

```
  PUBMED   11206551
REFERENCE  2  (bases 1 to 13501)
  AUTHORS  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
           Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
           Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
           Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
           Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
           Welch,R.A. and Blattner,F.R.
  TITLE    Direct Submission
  JOURNAL  Submitted (22-OCT-2000) Laboratory of Genetics, University of
           Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
```

```
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match           100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity  100.0%;  Pred. No. 0.0013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db    9635 GTTGCTGGATCTGCTGTTTGAAGCG 9659


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
```

```
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
     gene       2781. .3686
                /gene="ECs0755"
     CDS        2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
     gene       complement(3720. .4322)
                /gene="ECs0756"
     CDS        complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
     gene       complement(4332. .5984)
                /gene="ECs0757"
     CDS        complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
     gene       complement(6092. .7372)
                /gene="ECs0758"
     CDS        complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
     gene       complement(7533. .7850)
                /gene="ECs0759"
     CDS        complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"

  Query Match             100.0%;  Score 25;  DB 2;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.00097;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db  92367 GTTGCTGGATCTGCTGTTTGAAGCG 92391


RESULT  16
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES            Location/Qualifiers
     source         1. .1121
```

```
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT       289 a    296 c    325 g    211 t
ORIGIN


  Query Match             96.0%;  Score 24;  DB 10;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.0055;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagc 24
           |||||||||||||||||||||||||
Db      98 GTTGCTGGATCTGCTGTTTGAAGC 121


RESULT  17
AE004192
LOCUS       AE004192    12891 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 100 of 251 of the complete
            chromosome.
ACCESSION   AE004192 AE003852
VERSION     AE004192.1  GI:9655581
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12891
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(84. .1370)
                     /gene="VC1111"
     CDS             complement(84. .1370)
                     /gene="VC1111"
                     /note="similar to GB:J04423 SP:P12995 PID:457106
                     PID:490217 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adenosylmethionine-8-amino-7-oxononanoate
                     aminotransferase"
                     /protein_id="AAF94270.1"
                     /db_xref="GI:9655582"
                     /translation="MMMDLAFDRCHIWHPYTSTLTPLTCYPVSRADGVMITLEDGRSL
                     VDGMSSWWAAIHGYNHPKLNQAAHQQIEQMSHIMFGGLTHQPAIELCQKLLKLAPNNL
                     QQVFLADSGSVAVEVSLKMALQYWHSKGQPRAKFLTLRHGYHGDTFAAMSVTDPDNSM
                     HSLYKGFLPEHIFANSPEGGFFDAWDERDIADFRHKLTEHHHQIAAVILEPIVQGAGG
                     MRIYHPEFLRQVRALCDEFGVLLILDEIATGFGRTGKLFACEHAGIQPDILCVGKALT
                     GGYMTLSATLTTQQVADTVCAGEAGCFMHGPTFMGNPLACAVASASLSLIEQGDWQQQ
                     VANIEAFFAEQLPKLNDSPRVKQTRWLGAIGVVETHLPVNMEVIQALFVEHGVWIRPF
                     GRLIYLMPPYISQPEHLQKLVDAMAAALNQPDCFKQ"
     gene            1497. .2549
                     /gene="VC1112"
     CDS             1497. .2549
                     /gene="VC1112"
                     /note="similar to GB:J04423 SP:P12996 PID:145425
                     PID:490219 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
                     /protein_id="AAF94271.1"
                     /db_xref="GI:9655583"
                     /translation="MEVRHNWTVAEVKALLDKPFMDLLFEAQQVHRLHHPHNHVQVST
                     LLSIKTGACPEDCKYCPQSAHYRTDVDKERLMEVERVLDAAQKAKNSGSTRFCMGAAW
                     KNPKERDMPLLKEMIRGVKDMGLETCMTLGMLTPDQAQQLAQAGLDYYNHNLDTSPEF
                     YGNIITTRTYQDRLDTLSHVRDAGMKICSGGIIGMGESTNDRAGLLVELANLPTHPES
                     VPINMLVKVKGTPLEQVDDVEPFDFVRLIAVARIMMPKSAVRLSAGREKMNEQMQALC
                     FMAGANSIFYGCKLLTTPNPAEDSDMLLFKKLGINREQVAQKPDEITENELLDRVVER
                     VAARPTASDLFYDAAL"
     gene            2533. .3687
                     /gene="VC1113"
     CDS             2533. .3687
                     /gene="VC1113"
                     /note="similar to GB:J04423 SP:P12998 PID:145426
                     PID:490225 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAF94272.1"
                     /db_xref="GI:9655584"
                     /translation="MTLRFKSRIEQALNEREHSRLTRRLQLLDRSAQGELIQQDRTYL
                     NFSSNDYLGLANDAELVQAWQTGLARFGAGSGASPMVTGFSSAHAELEHELCHWLGYE
                     RAVLFSSGFSANQALLFTLLEKEDLLLQDKLNHASLMEAGMLSPATMKRFKHNDTEHL
                     RQLLHEQSNSLVVTEGVFSMDGDCAPLAQIATLVKQRAWLMVDDAHGIGVLGEDGAGS
                     CQAAGIHPEILVVTFGKAFGLAGAAVLCDAHVGDYLTQFARHHVYSTAMPPAQAHALT
                     HAARMIQSQSWRREQLAELLACFDEQCRNIPGFVATQTPIKPWLLGSSDSALSASHEL
                     KQQGIWVSAIRPPTVPVGSARLRITITAAHTQAQIRRLSEQLKHVMEKDT"
     gene            3549. .4487
                     /gene="VC1114"
     CDS             3549. .4487
                     /gene="VC1114"
                     /note="similar to GB:J04423 SP:P12999 PID:145427
                     PID:490221 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis protein BioC"
                     /protein_id="AAF94273.1"
                     /db_xref="GI:9655585"
                     /translation="MGECHSTTHRACRQCAFTYHDNGGAYPSADTPLIGAVKTCDGEG
                     HMSMTATELCVELKDKSAIAQAFGKAAAHYDQHAAFQRDVGLRLLQKMPSCLKGLRVL
                     DLGCGTGYFSALLRERGAQVVCADISHAMLEQAKQRCGDEGMSYQLADAEQLPFASAC
                     FDMVFSSLALQWCEDLSLPLSEIRRVLKPHGQAFLSTLLDGSLFELEQAWRSVDHHRH
                     INQFISINQVKIALAQAGCSQHHLDLAAITVWYEAAFMLMRDLKGIGANHVSGRSTGL
                     ISRRTLAKVELAYQSFRNQQGLVPATYQVCLGVIHR"
     gene            4463. .5179
                     /gene="VC1115"
     CDS             4463. .5179
                     /gene="VC1115"
                     /note="similar to GB:S68059 SP:P13000 PID:490223
                     PID:145428 PID:460901; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAF94274.1"
                     /db_xref="GI:9655586"
                     /translation="MFRGYSPMRNAIFIAGTDTDVGKTVASKAILQALATHNIATIGY
```

```
                KPVAAGSDKTEFGYRNSDALHLMKAATVDMPYEDVNPYALVLPTSPHIAAKHENVTID
                YALLSNKLSKHKQNAELVVVEGAGGWRVPTSDSDCLSTWVKQERLPVILVVGIKLGCL
                SHAILTAEAIRADGLELVGWIANRINPGTEHYAEIIEHLEGRLGTPKLGEIPYMPKAK
                RQELGKFIQLDHLLEPDTVA"
     gene       complement(5261. .5470)
                /gene="VC1116"
     CDS        complement(5261. .5470)
                /gene="VC1116"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF94275.1"
                /db_xref="GI:9655587"
                /translation="MSRRNLTTNKGKHMAEAHSYQMAIDLLCCHLGISEQEARTQLGL
                EAKEQTLQQQIVDTQHSLMGLSAEK"
     gene       5697. .6560
                /gene="VC1117"
     CDS        5697. .6560
                /gene="VC1117"
                /note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
                PID:1736470; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="heat shock protein HtpX"
                /protein_id="AAF94276.1"
                /db_xref="GI:9655588"
                /translation="MKRILLFLATNLAVVLVLSVVLNIVYAVTGMQPGSLSGLLVMAA
                VFGFGGAFISLLMSKSMALRSVGGVVIDTPRNEMEHWLLETVRRQANQAGIGMPTVAI
                YDAPDMNAFATGAKRDDSLVAVSTGLLHNMTRDEAEAVLAHEVSHIANGDMVTMTLMQ
                GVVNTFVIFLSRFIANIVASRDSEEGEGSNMMVYFGVSMVLELVFGFLASFITMWYSR
                HREFHADAGAAQLVGKHKMIAALERLKMGQESHLEGSMMAFGITGKRSLSELMMTHPP
                LEKRIAALRNM"
     gene       6770. .7198
                /gene="VC1118"
     CDS        6770. .7198
                /gene="VC1118"
                /note="similar to PID:1653608; identified by sequence
                similarity; putative".
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, putative"
                /protein_id="AAF94277.1"
                /db_xref="GI:9655589"
                /translation="MSIMDVQHVAQFYQQLDKSQLHRLIEIYHPDVVFEDAAHRIEGF
                DALYQYFLNLYQNVHTCTFTIHEQYAVNEGAFLVWTMHLRHPKLAKGEQVDVKGVSHL
                HFAEGKVTYHRDYFDMGEMLYEQLPVLGQVIRAIKRRLGQ"
     gene       7195. .7920
                /gene="VC1119"
     CDS        7195. .7920
                /gene="VC1119"
                /note="similar to SP:P25970; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="oxidoreductase, short-chain
                dehydrogenase/reductase family"
                /protein_id="AAF94278.1"
                /db_xref="GI:9655590"
                /translation="MTAVFITGATSGIGKQLALDYAKQGWQVIACGRNQSVLDSLHTQ
                YANIFPLAFDVTDHPGTKAALAQLPCQPELWILNAGDCEYIDDGKMDVTLMARVFNIN
                VLGVAAVIEGIQPHLSCGHRVAIVGSIASELALPRAEAYGASKAAVAYLARTLQLDWR
                PLGIEVTTIFPGFVATPLTDRNTFAMPMIITVERAAQEIKAGLARGVSQLYFPKRFTW
                LIRLLGALPYAWQGRLVRRLLKA"
     gene       7931. .9211
                /gene="VC1120"
     CDS        7931. .9211
                /gene="VC1120"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11

Query Match             80.0%;  Score 20;  DB 1;  Length 12891;
  Best Local Similarity   100.0%;  Pred. No. 0.65;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 tggatctgctgtttgaagcg 25
          ||||||||||||||||||||
Db   1558 TGGATCTGCTGTTTGAAGCG 1577


RESULT  18
S64534/c
LOCUS       S64534         383 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc4 allele, partial cds.
ACCESSION   S64534
VERSION     S64534.1  GI:407575
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136269] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .383
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>383
                     /gene="dec-1"
                     /allele="Fc4"
     CDS             <1. .>383
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27805.2"
                     /db_xref="GI:7548373"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQHDHQMAQQMAQQGLMMTEQRQRQWSEDQAKI
                     QQAQQMAQQTPMMMPQMQQRQWTEDP"
BASE COUNT      143 a     90 c    102 g     48 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 96;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    159 TTGCTGGATCTGCTGTT 143


RESULT  19
S64532/c
LOCUS       S64532         461 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc3 allele, partial cds.
ACCESSION   S64532
VERSION     S64532.1  GI:407573
```

```
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136268] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>461
                     /gene="dec-1"
                     /allele="Fc3"
     CDS             <1. .>461
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27804.2"
                     /db_xref="GI:7537267"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMVMQQRQWSEEQAKIQHDQQM
                     AQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQTPMMMPQMHQRQWTEDP"
BASE COUNT      174 a    108 c    122 g     57 t
ORIGIN


  Query Match           68.0%;  Score 17;  DB 96;  Length 461;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    237 TTGCTGGATCTGCTGTT 221


RESULT  20
DMDECG1B/c
LOCUS       DMDECG1B      516 bp    DNA             INV       02-DEC-1993
DEFINITION  D.melanogaster dec-1 gene 3'.
ACCESSION   X76231
VERSION     X76231.1  GI:433551
KEYWORDS    dec-1 gene.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. I. Restriction
            site mapping and limited sequence comparison in the melanogaster
            species subgroup
  JOURNAL   J. Mol. Evol. 33 (4), 321-331 (1991)
  MEDLINE   92130257
COMMENT     X76230, X76231, X76232, x76233 are related accession numbers.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C3-4"
     gene            1. .516
                     /gene="dec-1"
     exon            <1. .>516
                     /gene="dec-1"
                     /number=2
     CDS             <1. .>516
                     /gene="dec-1"
                     /codon_start=1
                     /product="fc proteins"
                     /protein_id="CAA53816.1"
                     /db_xref="GI:433552"
                     /db_xref="FLYBASE:FBgn0000427"
                     /db_xref="SWISS-PROT:P18170"
                     /translation="KEAQRRARNSGINTQKANALKRQAKSQDQTLSKEDIVQIMAYAY
                     RMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQ
                     RQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEE
                     QAKIQQNQQQIQ"
     repeat_unit     205. .282
                     /gene="dec-1"
     repeat_unit     283. .360
                     /gene="dec-1"
     repeat_unit     361. .438
                     /gene="dec-1"
     repeat_unit     439. .516
                     /gene="dec-1"
BASE COUNT      194 a    120 c    137 g     65 t
ORIGIN


  Query Match           68.0%;  Score 17;  DB 6;  Length 516;
  Best Local Similarity  100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    441 TTGCTGGATCTGCTGTT 425


RESULT  21
S64530/c
LOCUS       S64530        539 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc2 allele, partial cds.
ACCESSION   S64530
VERSION     S64530.1  GI:407569
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136266] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /note="variant Israel"
     gene            <1. .>539
```

```
                     /gene="dec-1"
                     /allele="Fc2"
     CDS             <1. .>539
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27802.2"
                     /db_xref="GI:7537265"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQ
                     IQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQM
                     AQQTPMMMPQMQQRQWTEDP"
BASE COUNT      208 a    124 c    141 g     66 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 96;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    315 TTGCTGGATCTGCTGTT 299


RESULT  22
S64531/c
LOCUS       S64531        539 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc2 allele, partial cds.
ACCESSION   S64531
VERSION     S64531.1  GI:407571
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136267] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>539
                     /gene="dec-1"
                     /allele="Fc2"
     CDS             <1. .>539
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27803.2"
                     /db_xref="GI:7537266"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQ
                     IQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQM
                     AQQTPMMMPQMQQRQWTEDP"
BASE COUNT      209 a    124 c    140 g     66 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 96;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    315 TTGCTGGATCTGCTGTT 299


RESULT  23
S64529/c
LOCUS       S64529        617 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc1 allele, partial cds.
ACCESSION   S64529
VERSION     S64529.1  GI:407567
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136265] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>617
                     /gene="dec-1"
                     /allele="Fc1"
     CDS             <1. .>617
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27801.2"
                     /db_xref="GI:7537264"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQ
                     IQQNPMMVQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQ
                     QGLMMTEQRQRQWSEDQAKIQQAQQMAQQTPMMMPQMQQRQWTEDP"
BASE COUNT      239 a    143 c    160 g     75 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 96;  Length 617;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    393 TTGCTGGATCTGCTGTT 377
```

```
RESULT  24
DSDEC1TR6/c
LOCUS       DSDEC1TR6     774 bp    DNA             INV       25-SEP-1996
DEFINITION  D.sechellia DNA from repeated region of dec-1 locus.
ACCESSION   X96931
VERSION     X96931.1  GI:1262763
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila sechellia.
  ORGANISM  Drosophila sechellia
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 774)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Drosophila sechellia"
                     /db_xref="taxon:7238"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .774
                     /rpt_type=TANDEM
     gene            1. .774
                     /gene="dec-1"
     exon            <1. .>774
                     /gene="dec-1"
                     /number=2
BASE COUNT      293 a    178 c    213 g     90 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 774;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db     81 TTGCTGGATCTGCTGTT 65


RESULT  25
DSDEC1TR2/c
LOCUS       DSDEC1TR2     852 bp    DNA             INV       25-SEP-1996
DEFINITION  D.simulans DNA from repeated region of dec-1 locus, strain
            Yugoslavia.
ACCESSION   X96927
VERSION     X96927.1  GI:1262760
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila simulans.
  ORGANISM  Drosophila simulans
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 852)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Drosophila simulans"
                     /strain="Yugoslavia"
                     /db_xref="taxon:7240"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .852
                     /rpt_type=TANDEM
     gene            1. .852
                     /gene="dec-1"
     exon            <1. .>852
                     /gene="dec-1"
                     /number=2
BASE COUNT      324 a    195 c    232 g    101 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 852;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          |||||||||||||||||
Db    237 TTGCTGGATCTGCTGTT 221


RESULT  26
DSDEC1TR3/c
LOCUS       DSDEC1TR3     852 bp    DNA             INV       25-SEP-1996
DEFINITION  D.simulans DNA from repeated region of dec-1 locus, strain
            Barcelona.
ACCESSION   X96928
VERSION     X96928.1  GI:1262761
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila simulans.
  ORGANISM  Drosophila simulans
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 852)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Drosophila simulans"
                     /strain="Barcelona"
                     /db_xref="taxon:7240"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .852
                     /rpt_type=TANDEM
     gene            1. .852
                     /gene="dec-1"
     exon            <1. .>852
                     /gene="dec-1"
                     /number=2
BASE COUNT      324 a    195 c    232 g    101 t
```

```
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 852;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db    237 TTGCTGGATCTGCTGTT 221


RESULT  27
DMDEC1TR1/c
LOCUS       DMDEC1TR1     930 bp    DNA             INV       25-SEP-1996
DEFINITION  D.melanogaster DNA from repeated region of dec-1 locus, strain
            Samarkand.
ACCESSION   X96926
VERSION     X96926.1  GI:1262758
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 930)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Drosophila melanogaster"
                     /strain="Samarkand"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .930
                     /rpt_type=TANDEM
     gene            1. .930
                     /gene="dec-1"
     exon            <1. .>930
                     /gene="dec-1"
                     /number=2
BASE COUNT      352 a    213 c    254 g    111 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 930;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db    393 TTGCTGGATCTGCTGTT 377


RESULT  28
DRODEC1A/c
LOCUS       DRODEC1A     3741 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fc125 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35887 M35688
VERSION     M35887.2  GI:11386620
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 3741)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157181.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .3741
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .3741
                     /gene="dec-1"
     CDS             76. .3447
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fc125 protein precursor"
                     /protein_id="AAA28446.1"
                     /db_xref="GI:157182"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARTINPKQPGEVGGSESQKSNSNPPTTLTP
                     APQEQPQEHRVHKSPSSAPSETEIENAPSSDPQVGSIFTYGEGLLHPFMGLLPVERPD
                     DPWNQKPYDPHHPLYTGGGSYDAYLRDGRHRRDTHIMGQGTQHGILTPGMLERLLRIK
                     MDFQRRFPPFIQGYAESPYESYPCRGTASGSGQDF"
     sig_peptide     76. .144
                     /gene="dec-1"
     mat_peptide     145. .3444
                     /gene="dec-1"
                     /product="defective chorion-1 fc125 protein"
BASE COUNT     1091 a    987 c    993 g    670 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 3741;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db   1944 TTGCTGGATCTGCTGTT 1928


RESULT  29
DRODEC1B/c
LOCUS       DRODEC1B     3755 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fc106 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35888 M35688
VERSION     M35888.2  GI:11386621
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 3755)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157183.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .3755
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .3755
                     /gene="dec-1"
     CDS             76. .2928
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fc106 protein precursor"
                     /protein_id="AAA28447.1"
                     /db_xref="GI:157184"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARATHFRQ"
     sig_peptide     76. .144
                     /gene="dec-1"
     mat_peptide     145. .2925
                     /gene="dec-1"
                     /product="defective chorion-1 fc106 protein"
BASE COUNT     1094 a    991 c    997 g    673 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 3755;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db   1944 TTGCTGGATCTGCTGTT 1928


RESULT  30
DRODEC1C/c
LOCUS       DRODEC1C     5697 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fc177 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35889 M35688
VERSION     M35889.2  GI:11386622
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 5697)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157185.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .5697
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .5697
                     /gene="dec-1"
     CDS             76. .4845
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fc177 protein precursor"
                     /protein_id="AAA28448.1"
                     /db_xref="GI:157186"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
```

```
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARTINPKQPGEVGGSESQKSNSNPPTTLTP
                     APQEQPQEHRVHKRLAHFHRFGQEAGLNATTSKGCGCGRLDCLCGRSCRCGRRGLESR
                     VVSSRTSGTCQCKASHRNKRCEYGTLETIDEGSLNELRREYKLGLKEITLSPDEDPAE
                     ALMRYNAASIREALERASMEPLEIGGDQYEEDAQQEPMEEEQLQHDPNTEPQYNHKDF
                     VRLTTSTASPITSTTEAATPTGSDSTSEATVTPEVTTTTSTSTTTTTESTKDEGLDMQ
                     QDSQAEAESSHVTKSISKQEAXIHQLHSIVEELKNEILKLNLRCSTIISNNVAKEPVT
                     EKNPPVVEEPSKQEDKPKVEEKVIAEEQAPVEPEEDLEEDEDSTSISTTTETPSPSGS
                     YSTKPGLSLGSPRVDEQSGSSNKLDYDDDNNWQRILANRGYDTDYLTKSHERQFAQGQ
                     NLEMPKNCNYDGNGSQEYGPYPEFQADEPSTDTEGKAKRALSVKQQAQLLNAALNDSG
                     SDSSDGTTTTTTPSPYAMRGKFVRRRSTARRVPIPKIGKASDEVWVRSPRQAKMPQRP
                     KKSMSKPKKQSSQVTTQATVSSTKLDSLVDVLKDLVRLQIQKEKKSSLLRTQSNNLSK
                     TKPKSIKPVKVIKRKRLRRRQHKSIATTIRSPIQTKA"
     sig_peptide     76. .144
                     /gene="dec-1"
     mat_peptide     145. .4842
                     /gene="dec-1"
                     /product="defective chorion-1 fc177 protein"
BASE COUNT     1704 a   1469 c   1485 g   1038 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 6;  Length 5697;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||||
Db   1944 TTGCTGGATCTGCTGTT 1928


RESULT  31
AC017995/c
LOCUS       AC017995    33204 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017995
VERSION     AC017995.1  GI:6553195
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 33204)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212937 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .33204
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     8981 a   7720 c   7703 g   8800 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 65;  Length 33204;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||||
Db  15905 TTGCTGGATCTGCTGTT 15889


RESULT  32
AC016952/c
LOCUS       AC016952   161540 bp    DNA             HTG       27-AUG-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-220O14, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC016952
VERSION     AC016952.11  GI:9929556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161540)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161540)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9719634.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMUN
                Center clone name: RP11-220O14
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 138744 bases at least Q40
                Consensus quality: 149435 bases at least Q30
                Consensus quality: 153457 bases at least Q20
                Estimated insert size: 155950; sum-of-contigs estimation
                Estimated insert size: 211269; agarose-fp estimation
                Quality coverage: 3.5x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5164: contig of 5164 bp in length
            *    5165     5264: gap of unknown length
            *    5265    55000: contig of 49736 bp in length
            *   55001    55100: gap of unknown length
            *   55101    84235: contig of 29135 bp in length
            *   84236    84335: gap of unknown length
            *   84336   104093: contig of 19758 bp in length
            *  104094   104193: gap of unknown length
            *  104194   117951: contig of 13758 bp in length
            *  117952   118051: gap of unknown length
            *  118052   126732: contig of 8681 bp in length
            *  126733   126832: gap of unknown length
            *  126833   135507: contig of 8675 bp in length
            *  135508   135607: gap of unknown length
            *  135608   141357: contig of 5750 bp in length
            *  141358   141457: gap of unknown length
            *  141458   145691: contig of 4234 bp in length
            *  145692   145791: gap of unknown length
            *  145792   149739: contig of 3948 bp in length
            *  149740   149839: gap of unknown length
            *  149840   152725: contig of 2886 bp in length
            *  152726   152825: gap of unknown length
            *  152826   155298: contig of 2473 bp in length
            *  155299   155398: gap of unknown length
            *  155399   158289: contig of 2891 bp in length
            *  158290   158389: gap of unknown length
            *  158390   160296: contig of 1907 bp in length
            *  160297   160396: gap of unknown length
            *  160397   161540: contig of 1144 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-220014"
BASE COUNT    53135 a  28040 c  27770 g  51176 t   1419 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 64;  Length 161540;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgttt 19
          |||||||||||||||||
Db  22839 TGCTGGATCTGCTGTTT 22823


RESULT  33
AE003442
LOCUS       AE003442   306267 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386054 section 26
            of 35, complete sequence.
ACCESSION   AE003442 AE002566
VERSION     AE003442.1  GI:7290831
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 306267)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 306267)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .306267
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(32686. .32823,32937. .33134,33225. .33387,34597. .34762,
                     35618. .35749)
                     /gene="CG1402"
                     /note="Partial transcript."
                     /product="CT3206"
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
     gene            <32686. .>35749
                     /gene="CG1402"
                     /map="7C3-7C4"
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
     CDS             join(33264. .33387,34597. .34762,35618. .35639)
                     /gene="CG1402"
                     /note="CG1402 gene product; Partial transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
                     /protein_id="AAF46274.1"
                     /db_xref="GI:7290832"
                     /translation="MSEAQHKSQGIVGLSLMVQIGETPNPELRIITSTFNKVLYRGFS
                     TPIRHISVRSLLPNTDHYITYEGSTTHPGCWESTVWIIVNKPIYITKQEVSLCCTNSD
                     D"
     mRNA            complement(48924. .>50117)
                     /gene="CG10920"
                     /product="CT30589"
                     /db_xref="FLYBASE:FBan0010920"
                     /db_xref="FLYBASE:FBgn0029963"
```

```
gene            complement(<48924. .>50117)
                /gene="CG10920"
                /map="7C3-7C5"
                /db_xref="FLYBASE:FBan0010920"
                /db_xref="FLYBASE:FBgn0029963"
CDS             complement(48924. .50033)
                /gene="CG10920"
                /note="CG10920 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0010920"
                /db_xref="FLYBASE:FBgn0029963"
                /protein_id="AAF46275.1"
                /db_xref="GI:7290833"
                /translation="MRHPTPPPLAVQHARSGAMVVELAVDPSRAMRSTNQQQPIDRDQ
                LRISTEESCQGSELVVASSHLSKKIQEHRRPNQLVRFCLRLGYNTLLYPYEMAKVLIQ
                LGHEPLQAKPFVMRLIQRRPRLFLPSVHRYVQHIQHIDGYTGMYRGLTARLAASIVDY
                LLGDLLLATLHFAPYKRGPKEGLSLKEFWWNLTRNSLRITTVVVITHPFYVVMVRQIA
                QFVGGEHVYEGLVGSLMTLAQQEGCAGLFAGMVPRLLGEWSVLFITSALSHLCRRLLP
                MSELQHQYNTAVIQMMASLVAYPLEVTSTCMAGTGAPLTACEPPSMPLYNHWVDCLTD
                LYARGGQNRGAILFWRTVPKIQLLRSKDNYTVRSI"
mRNA            <54229. .>54681
                /gene="CG1409"
                /product="CT3292"
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
gene            <54229. .>54681
                /gene="CG1409"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
CDS             54229. .54681
                /gene="CG1409"
                /note="CG1409 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
                /protein_id="AAF46276.1"
                /db_xref="GI:7290834"
                /translation="MARYLAQIIILGAQLVGRALVKTMRQELQAFEDAARLQETLKAN
                DPNSGRSAVAKTMTLAEAQQILDVSDLTNRQAIDTHYQHLFRVNDKSTGGSFYIQSKV
                FRAKERIDQELERTELLVKTDDSHSILPTPPESSQFQCENKEPGQKSR"
mRNA            complement(join(<60589. .60716,60988. .62089,62143. .>62916))
                /gene="CG15326"
                /product="CT35326"
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
gene            complement(<60589. .>62916)
                /gene="CG15326"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
CDS             complement(join(60589. .60716,60988. .62089,62143. .62916))
                /gene="CG15326"
                /note="CG15326 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
                /protein_id="AAF46277.1"
                /db_xref="GI:7290835"
                /translation="MESSSDWDLAEALAQVVANSEMGRFKTLYIYTHTNSQSTGGHLE
                ELLDQVLMIVPNNLQARRLLLQQSMEYKPYVHAVLALVDGLPSLSAIYARIRATQDLS
                HTLIYMSMPTDAYGEEMQATLRFLWRLSVLNVGVVLRPPGDHILMVSYFPFSALHGCQ
                VISANVVNRYQVGTKRWASQDYFPSKLGNFYGCLLTCATWEDMPYLVWRPDGSGSFVG
                IEGALLQFMAENLNFTVGLYWMNKEEVLATFDESGRIFDEIFGHHADFSLGGFHFKPS
                AGSEIPYSQSTYYFMSHIMLVTNLQSAYSAYEKLSFPFTPLLWRAIGLVLILACLLLM
                LLVRWRHHHELPRNPYYELLVLTMGGNLEDRWVPQRFPSRLVLLTWLFATLVLRSGYQ
                SGMYQLLRQDTQRNPPQTISEVLAQHFTIQLAEVNEARILASLPELRPEQLVYLEGSE
                LQSFPALAQQSGSSARVAILTPYEYFGYFRKVHPMSRRLHLVRERIYTQQLAFYVRRH
                'SHLVGVLNKQIQHAHTHGFLEHWTRQYVSAVDEKDESVARIASTSYSTLDGIDGDPSL
                SESEEDQQVAPVRQNVLSMRELAALFWLILWANLGAVVVFVLELLLPRIKLRKILRKM
                KKSTRASATTTSTLSSPSTTKDIPFSCKDGFQDSWPKCSLLVS"
mRNA            complement(join(63233. .63902,64113. .65786,65862. .65986,
                66185. .66447,66505. .68920,68994. .>69146))
                /gene="dec-1"
                /note="Nucleotide sequence of the Celera sequence differs
                from the published sequence for this transcript."
                /product="CT7110"
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
gene            complement(<63233. .>69146)
                /gene="dec-1"
                /note="CG2175"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
CDS             complement(join(63233. .63902,64113. .65786,65862. .65986,
                66185. .66447,66505. .68854))
                /gene="dec-1"
                /note="dec-1 gene product; Nucleotide sequence of the
                Celera sequence differs from the published sequence for
                this transcript"
                /codon_start=1
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
                /protein_id="AAF46278.1"
                /db_xref="GI:7290836"
                /translation="MMFPALGSLLKWGSLFPAYSILGAIPDNLQPTAAASKVVLVLAD
                DATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVGSFLAQLPAMPTVPGL
                LGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILGQAALQNAFTFFNPAN
                FDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGTDAQASDISEVRVRPE
                DPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTTQNQDATEEKTLEDLR
                VEAKLRAFERQVIAELRMLQKIELMAKQMRSSAAAQNGDSPYRISYPLSRTPIHKITR
                ADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKSQDQTLSKEDIVQIMA
                YAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMM
                MQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQW
                SEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAK
```

```
  Query Match             68.0%;  Score 17;  DB 4;  Length 306267;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           ||||||||||||||||
Db   67205 TTGCTGGATCTGCTGTT 67221
```

```
RESULT  34
DMDEC1TR5/c
LOCUS       DMDEC1TR5     774 bp    DNA             INV       25-SEP-1996
DEFINITION  D.mauritiana DNA from repeated region of dec-1 locus.
ACCESSION   X96930
VERSION     X96930.1  GI:1262759
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila mauritiana.
  ORGANISM  Drosophila mauritiana
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 774)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .774
```

```
                     /organism="Drosophila mauritiana"
                     /db_xref="taxon:7226"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .774
                     /rpt_type=TANDEM
     gene            1. .774
                     /gene="dec-1"
     exon            <1. .>774
                     /gene="dec-1"
                     /number=2
BASE COUNT      290 a    182 c    212 g     90 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 6;  Length 774;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgt 17
          ||||||||||||||||
Db    159 TTGCTGGATCTGCTGT 144


RESULT  35
AF281080/c
LOCUS       AF281080     1614 bp    mRNA            VRT       01-OCT-2000
DEFINITION  Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
            complete cds.
ACCESSION   AF281080
VERSION     AF281080.1  GI:10442649
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A., Garriock,R.J., Li,D. and Krieg,P.A.
  TITLE     Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
  JOURNAL   Mech. Dev. 96 (2), 259-262 (2000)
  MEDLINE   20417824
REFERENCE   2  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A. and Krieg,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
            Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
            85724, USA
FEATURES             Location/Qualifiers
     source          1. .1614
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /note="expressed in ventral forebrain, lungbud primordia
                     and thyroid anlage"
     gene            1. .1614
                     /gene="Nkx2-1"
     CDS             162. .1205
                     /gene="Nkx2-1"
                     /note="thyroid transcription factor; TTF-1; T/ebp; chicken
                     and mouse Nkx2-1 ortholog; contains homeodomain"
                     /codon_start=1
                     /product="homeobox transcription factor Nkx2-1"
                     /protein_id="AAG17405.1"
                     /db_xref="GI:10442650"
                     /translation="MSMSPKHTTPFSVSDILSPLEESYKKVAMEGAGLGAPLTAAAYR
                     QSQVSQASMQQHGMGHNGPVSAAYHMTAAGVPQLSHTTMGGYCNGNLGNLGNMSELPP
                     YQDTMRNSSATGWYGANPDPRFSTISRFMGPSGGMNMGGLGNMGSLGDVGKSMTPLQA
                     TPRRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMIHLTPTQVKIWFQNHRYKM
                     KRQAKDKASQQQIQQDNSSCQQQQSPRRVAVPVLVKDGKPCQAGSNTPTAALQSHQQQ
                     TATAITITTNGLGPHQSHQTNSAGQSPDLVPHSNSPSSLQNQVTSLSHLNSSSSDYGS
                     AMSCSTLLYGRTW"
BASE COUNT      475 a    407 c    399 g    333 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 8;  Length 1614;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 tctgctgtttgaagcg 25
          ||||||||||||||||
Db    717 TCTGCTGTTTGAAGCG 702


RESULT  36
AF143940
LOCUS       AF143940     5000 bp    DNA             PLN       11-MAY-1999
DEFINITION  Arabidopsis thaliana SWI2/SNF2-like protein (DDM1) gene, complete
            cds; and tRNA-Glu gene, complete sequence.
ACCESSION   AF143940
VERSION     AF143940.1  GI:4726078
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 5000)
  AUTHORS   Jeddeloh,J.A., Stokes,T.L. and Richards,E.J.
  TITLE     Maintenance of genomic methylation requires a SWI2/SNF2-like
            protein
  JOURNAL   Nat. Genet. 22 (1), 94-97 (1999)
  MEDLINE   99251588
REFERENCE   2  (bases 1 to 5000)
  AUTHORS   Jeddeloh,J.A. and Richards,E.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1999) Biology, Washington University, One
            Brookings Drive, St. Louis, MO 63130, USA
FEATURES             Location/Qualifiers
     source          1. .5000
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     mRNA            join(<535. .566,772. .850,986. .1252,1354. .1440,1549. .1895,
                     1976. .2165,2251. .2426,2559. .2625,2703. .2892,2975. .3070,
                     3148. .3242,3317. .3436,3540. .3659,3745. .3843,3934. .4038,
                     4130. .>4354)
                     /gene="DDM1"
                     /product="SWI2/SNF2-like protein"
     gene            <535. .>4354
                     /gene="DDM1"
     CDS             join(535. .566,772. .850,986. .1252,1354. .1440,1549. .1895,
                     1976. .2165,2251. .2426,2559. .2625,2703. .2892,2975. .3070,
                     3148. .3242,3317. .3436,3540. .3659,3745. .3843,3934. .4038,
                     4130. .4354)
                     /gene="DDM1"
                     /codon_start=1
                     /product="SWI2/SNF2-like protein"
                     /protein_id="AAD28303.1"
                     /db_xref="GI:4726079"
                     /translation="MVSLRSRKVIPASEMVSDGKTEKDASGDSPTSVLNEEENCEEKS
                     VTVVEEEILLAKNGDSSLISEAMAQEEEQLLKLREDEEKANNAGSAVAPNLNETQFTK
                     LDELLTQTQLYSEFLLEKMEDITINGIESESQKAEPEKTGRGRKRKAASQYNNTKAKR
                     AVAAMISRSKEDGETINSDLTEEETVIKLQNELCPLLTGGQLKSYQLKGVKWLISLWQ
                     NGLNGILADQMGLGKTIQTIGFLSHLKGNGLDGPYLVIAPLSTLSNWFNEIARFTPSI
                     NAIIYHGDKNQRDELRRKHMPKTVGPKFPIVITSYEVAMNDAKRILRHYPWKYVVIDE
                     GHRLKNHKCKLLRELKHLKMDNKLLLTGTPLQNNLSELWSLLNFILPDIFTSHDEFES
                     WFDFSEKNKNEATKEEEEKRRAQVVSKLHGILRPFILRRMKCDVELSLPRKKEIIMYA
                     TMTDHQKKFQEHLVNNTLEAHLGENAIRGQGWKGKLNNLVIQLRKNCNHPDLLQGQID
                     GSYLYPPVEEIVGQCGKFRLLERLLVRLFANNHKVLIFSQWTKLLDIMDYYFSEKGFE
                     VCRIDGSVKLDERRRQIKDFSDEKSSCSIFLLSTRAGGLGINLTAADTCILYDSDWNP
                     QMDLQAMDRCHRIGQTKPVHVYRLSTAQSIETRVLKRAYSKLKLEHVVIGQGQFHQER
                     AKSSTPLEEEDILALLKEDETAEDKLIQTDISDADLDRLLDRSDLTITAPGETQAAEA
                     FPVKGPGWEVVLPSSGGMLSSLNS"
     variation       785
```

```
                     /gene="DDM1"
                     /note="site of ddm1-5 (som8) mutation; replace with 83 bp"
                     /replace=""
     variation       2384. .2385
                     /gene="DDM1"
                     /note="site of ddm1-6 (som4) mutation"
                     /replace=""
     misc_feature    3186
                     /gene="DDM1"
                     /note="alternate splice donor site used in ddm1-2"
     variation       3243
                     /gene="DDM1"
                     /note="site of ddm1-2 mutation"
                     /replace="a"
     variation       3337
                     /gene="DDM1"
                     /note="site of ddm1-7 (som5) mutation"
                     /replace="a"
     tRNA            complement(4755. .4826)
                     /product="tRNA-Glu"
                     /evidence=not_experimental
BASE COUNT     1386 a    909 c   1029 g   1676 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 5000;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgtt 18
          ||||||||||||||||
Db   1132 TGCTGGATCTGCTGTT 1147


RESULT  37
YLI012084
LOCUS       YLI012084    6887 bp    DNA             PLN       23-SEP-1999
DEFINITION  Yarrowia lipolytica pex10 gene for peroxin 10.
ACCESSION   AJ012084
VERSION     AJ012084.1  GI:5921565
KEYWORDS    peroxin 10; pex10 gene.
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 6887)
  AUTHORS   Le Dall,M.T., Gaillardin,C. and Nicaud,J.M.J.
  TITLE     PEX10 is essential for peroxisome biogenesis in Yarrowia lipolytica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6887)
  AUTHORS   Nicaud,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
            Microorganismes, INRA-CNRS, BP 01, Tiverval-Grignon, F78850, FRANCE
FEATURES             Location/Qualifiers
     source          1. .6887
                     /organism="Yarrowia lipolytica"
                     /strain="W29"
                     /db_xref="taxon:4952"
     gene            complement(2293. .3357)
                     /gene="pex10"
     CDS             complement(2293. .3357)
                     /gene="pex10"
                     /function="essential for peroxisome biogenesis"
                     /codon_start=1
                     /evidence=experimental
                     /product="peroxin 10"
                     /protein_id="CAB56484.1"
                     /db_xref="GI:5921566"
                     /translation="MSDNTTIKKPIRPKPIRTERLPYAGAAEIIRANQKDHYFESVLE
                     QHLVTFLQKWKGVRFIHQYKEELETASKFAYLGLCTLVGSKTLGEEYTNLMYTIRDRT
                     ALPGVVRRFGYVLSNTLFPYLFVRYMGKLRAKLMREYPHLVEYDEDEPVPSPETWKER
                     VIKTFVNKFDKFTALEGFTAIHLAIFYVYGSYYQLSKRIWGMRYVFGHRLDKNEPRIG
                     YEMLGLLIFARFATSFVQTGREYLGALLEKSVEKEAGEKEDEKEAVVPKKKSSIPFIE
                     DTEGETEDKIDLEDPRQLKFIPEASRACTLCLSYISAPACTPCGHFFCWDCISEWVRE
                     KPECPLCRQGVREQNLLPIR"
BASE COUNT     1793 a   1841 c   1485 g   1768 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 15;  Length 6887;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gatctgctgtttgaag 23
          ||||||||||||||||
Db   4051 GATCTGCTGTTTGAAG 4066


RESULT  38
AE004322/c
LOCUS       AE004322    11328 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 230 of 251 of the complete
            chromosome.
ACCESSION   AE004322 AE003852
VERSION     AE004322.1  GI:9657104
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 11328)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 11328)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11328
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(95. .286)
                     /gene="VC2521"
     CDS             complement(95. .286)
                     /gene="VC2521"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF95663.1"
                     /db_xref="GI:9657105"
                     /translation="MAANFDVLGIKREVEVPINYPYPLAILPVCHSKRQPLKLVPRLS
                     SACYRLCWMKTFIGFLDAK"
     gene            524. .1522
                     /gene="VC2522"
     CDS             524. .1522
                     /gene="VC2522"
```

```
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95664.1"
                /db_xref="GI:9657106"
                /translation="MAFIFIYYVGINMLIEAVVFLFIGLAILVWGADKLVFGSAALAR
                NIGISPLVIGMTILAMGSSAPEMMVSATAALEGKTDTAVGNVLGSNIANIALILGITA
                IVKPLSVGSAVLRRELPMMIGVTLIAGAILWDNHLGFYEGILLIALFAAFILVMLQVS
                RKEKQNGDALLDEQESEIPVGVSNPKAAFWVVVGLILLPIGAGMLVDNAVVIAKYFGM
                SDLVIGLTIIAVGTSLPELAASLAGALKGEDDMAVGNIIGSNVFNILAVMGLPGLLNP
                SLLSPEAMSRDFWVMLGVSLLLVGVALGKKRQITRLEGALLLCSFIGYITYLLMNI"
gene            1572. .2552
                /gene="VC2523"
CDS             1572. .2552
                /gene="VC2523"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95665.1"
                /db_xref="GI:9657107"
                /translation="MRMPHSFDYQTVAKQVLATEIHALQQLEQYINDDFARACAMILA
                NQTGKVVVMGMGKSGHIGKKIAATLASTGTSAFFVHPGEASHGDLGMIERGDIVLAIS
                NSGESSEILALLPVLKRLSIRVISMTGNPNSNMAKLADIHLQITVPREACPLELAPTS
                STTATLVMGDALAVALMQARGFTAEDFALSHPGGALGRKLLLKLNDIMHSGDALPKVA
                PQALIRDALLEISQKGLGMTAIVDEQDTLLGIFTDGDLRRILDKRIDIHSTVIADVMT
                RQPTVAQPNLLAVEGLNLMQAKRINGLMLVENNKLVGALNMHDLLKAGVM"
gene            2552. .3109
                /gene="VC2524"
CDS             2552. .3109
                /gene="VC2524"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95666.1"
                /db_xref="GI:9657108"
                /translation="MSSTVSTLYGEVEPSLLEIAKQIKLLICDVDGVFSDGLIYMGNQ
                GEELKTFHTRDGYGVKALMNAGIEIAIITGRRSQIVENRMKALGISLIYQGQDDKVQA
                YYDICQKLAIAPEQTGYIGDDLIDWPVMEKVALRVCVADGHPLLAQRANYVTHIKGGH
                GAVREVCDLILQARNELDVHKGLSL"
gene            3106. .3669
                /gene="VC2525"
CDS             3106. .3669
                /gene="VC2525"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF95667.1"
                /db_xref="GI:9657109"
                /translation="MSLSRIVYVLLLFIASWSLYYLLGQEQDSKIQVAPNLELPMFSG
                ENLENISYDEQGIRNYVITSIHLDHYAKSGNTLFKAPILKVYREGTLQEWEITARRGI
                LSKDQVLTLYDDVLAKNLLPDSGFDTLTTSEMSIQLKSRDFWADKPVELRGPQFETHG
                QAMKGNFADHSAELYNQVQGRYETLTP"
gene            3605. .4144
                /gene="VC2527"
CDS             3605. .4144
                /gene="VC2527"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95668.1"
                /db_xref="GI:9657110"
                /translation="MPIIRLNFITKYKVDMKRSHLSLIACLLASTQAYALSTDREQPV
                YIDSDSQQLDMKSNRVTFLGDVKLKQGSININADKLIVIRNAENGKIEEIEGYGNVAT
                FSQLTDEGKTLYGEAKELYYKVRDDELVMINKAMLSQDDSEIRGSKIRYKISLQKLIA
                DSDGKGRVSTVLQPQAIQE"
gene            4147. .4872
                /gene="VC2528"
CDS             4147. .4872
                /gene="VC2528"
                /note="similar to PID:2317737; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /protein_id="AAF95669.1"
                /db_xref="GI:9657111"
                /translation="MAILKAQHLAKSYKKRKVVSDVSLQVESGQIVGLLGPNGAGKTT
                SFYMIVGLVARDEGTITIDDNDISILPMHSRSRMGIGYLPQEASIFRKLSVEDNIMAV
                LQTREELTHEERQDKLEDLLEEFHIQHIRKSAGMALSGGERRRVEIARALAANPQFIL
                LDEPFAGVDPISVIDIKKIIEHLRDRGLGVLITDHNVRETLDVCEKAYIVSQGRLIAE
                GTPQDVLNNEQVKQVYLGEQFRL"
gene            4928. .6391
                /gene="VC2529"
CDS             4928. .6391
                /gene="VC2529"
                /note="similar to PID:2317738; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase sigma-54 factor"
                /protein_id="AAF95670.1"
                /db_xref="GI:9657112"
                /translation="MKPSLQLKLGQQLAMTPQLQQAIRLLQLSTLDLQQEIQEALESN
                PLLEVEENQDEAPSLDSVRMTEESPREPEELYEPEPQDSSDLIEKSEISAELEMDTTW
                DEVYSANTGSTGLALDDDAPIYQGETTQTLQDYLHWQLDLTPFSDVDRTIAVALIDAI
                DDYGYLTVSLEEIQESLRSDDIELDEIEAVRKRIQQFDPFGVASLNLQDCLLLQLTTY
                PCDTPWLEEARLLLSQYIDDLGNRDYKTILKETKLKEEDLREILQLIQQLDPRPGSRI
                AQDHAEYVIPDVSVYKEQGRWLVTINPDSVPKLKINQQYADLMRGNNAESNYIRTNLQ
                EAKWLIKSLESRNETLLKVAKCIVEHQHDFFEYGEEAMKPMVLNDVAMAVEMHESTIS
                RVTTQKYMHTPRGIFELKYFFSSHVSTDNGGECSSTAIRALIKKLVAAENPAKPLSDS
                KIATLLADQGIQVARRTIAKYRESLGIAPSSQRKRLL"
gene            6416. .6703
                /gene="VC2530"
CDS             6416. .6703
                /gene="VC2530"
                /note="similar to PID:2317739; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="sigma-54 modulation protein, putative"
                /protein_id="AAF95671.1"
                /db_xref="GI:9657113"
                /translation="MQINIQGHHIDLTDSMQDYVHSKFDKLERFFDHINHVQVILRVE
                KLRQIAEATLHVNQAEIHAHADDENMYAAIDSLVDKLVRQLNKHKEKLSSH"
gene            6706. .7152
                /gene="VC2531"
CDS             6706. .7152
                /gene="VC2531"
                /note="similar to PID:2317740; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PTS system, nitrogen regulatory IIA component"
                /protein_id="AAF95672.1"
                /db_xref="GI:9657114"
                /translation="MQLSEVLSLDCTKSAVQCSSKKRALEIISEIAALHTGQNATELF
                ECMLSREKMGSTGIGNGIAIPHARMSDSENAVAVLLQCEEPIEFDSIDNRPVDLLFAL
                LVPEDQCKEHLKTLACMAERLNDKQILKQLRNAASDQELYNIMVNQ"
gene            7165. .8025
                /gene="VC2532"
CDS             7165. .8025
                /gene="VC2532"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95673.1"
                /db_xref="GI:9657115"
                /translation="MRLIVVSGQSGAGKSVALRVLEDLGYYCVDNLPVSLLTAFIQSV
                QGSQQNVAVSIDIRNLPKEPSLVQDVLDQLKQNNDVSMLFLDASKETLLKRYSETRRI
```

```
  Query Match             64.0%;  Score 16;  DB 1;  Length 11328;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgtttga 21
           ||||||||||||||||
Db    7838 TGGATCTGCTGTTTGA 7823


RESULT  39
AE004026
LOCUS       AE004026    13856 bp    DNA             BCT       06-JUL-2000
DEFINITION  Xylella fastidiosa, section 172 of 229 of the complete genome.
ACCESSION   AE004026 AE003849
VERSION     AE004026.1  GI:9107228
KEYWORDS    .
SOURCE      Xylella fastidiosa.
  ORGANISM  Xylella fastidiosa
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 13856)
  AUTHORS   Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
            Terenzi,M.F., Truffi,D., Tsai,S.M., Tsuhako,M.H., Vallada,H., Van
            Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
            Meidanis,J. and Setubai,J.C.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis, Sao Paulo, Brazil
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
REFERENCE   2  (bases 1 to 13856)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
  REMARK
FEATURES             Location/Qualifiers
     source          1. .13856
                     /organism="Xylella fastidiosa"
                     /db_xref="taxon:2371"
     gene            complement(195. .803)
                     /gene="XF2101"
     CDS             complement(195. .803)
                     /gene="XF2101"
                     /note="similar to GI|1652989 (percent identity: 30 %/query
                     alignment coverage:  87.6 %/subject alignment coverage:
                     67.6 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84900.1"
                     /db_xref="GI:9107229"
                     /translation="MSRCITGYRWWQWICVVVVLVFFWLVAVAGWIVWIGQRDQAASS
                     DVIIVLGAAAYDAKPSPVFEERIRHGLYLYQHGYAPTLIFTGGFGGSSARFAESQVAR
                     RYALRYDVPPRAILIETVSRTTRQNLIQARQLMRVYGLRRAIVVSDPLHMARALRLCR
                     ELDIDALASATPSTRFRSLHTRWRFLLQEVYFFHRDLVIQAN"
     gene            complement(800. .1042)
                     /gene="XF2102"
     CDS             complement(800. .1042)
                     /gene="XF2102"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84901.1"
                     /db_xref="GI:9107230"
                     /translation="MARVIMFYSSFALSVKSCIGVLTGMLLLHAYVFGCGLGGEFRAL
                     CCDVAWLCHVDISARLSGCGCSAGFFICGVSVLVVS"
     gene            1397. .1732
                     /gene="XF2103"
     CDS             1397. .1732
                     /gene="XF2103"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84902.1"
                     /db_xref="GI:9107231"
                     /translation="MKARLFLPILIMSAISGCATYTYSNGYYHGTPTVEYHYPPNYYY
                     PHGYGMHDLFNYPIPPIYPRDFYRTHIHLPYWQRPHHDPSRPLPPKQPIKQEQPNHIE
                     MGLRDKQPI"
     gene            1914. .2024
                     /gene="XF2104"
     CDS             1914. .2024
                     /gene="XF2104"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84903.1"
                     /db_xref="GI:9107232"
                     /translation="MNIADQDLMLLCSELELMKICPHCVDTLKHQRLAAR"
     gene            complement(2067. .2225)
                     /gene="XF2105"
     CDS             complement(2067. .2225)
                     /gene="XF2105"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84904.1"
                     /db_xref="GI:9107233"
                     /translation="MAAWSREEAWMEFYVFCVIFRNALFRVGTFYAVDVVAPAAALRL
                     QAQRSGLE"
     gene            2271. .4160
                     /gene="XF2106"
     CDS             2271. .4160
                     /gene="XF2106"
                     /note="similar to SP|P25756 (percent identity: 65 %/query
                     alignment coverage:  98.6 %/subject alignment coverage:
                     98.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose inhibited division protein A"
```

```
                     /protein_id="AAF84905.1"
                     /db_xref="GI:9107234"
                     /translation="MTDSFYGYDVIVIGAGHAGSEAALAAARTGAHTLLLTHNIETIG
                     AMSCNPAIGGIGKGHLVKEIDALGGAMAHAADAAGIQWRTLNASKGPAVRATRCQADR
                     ALYRAAIRQMIEGQARLNIFQAEVDDLLFEGDTVCGAITHTGLHFKAPAVVLTAGTFL
                     AGKIHIGPTQYAAGRMGDPPATMLAARLRERLFTVARLKTGTPPRIDGRTLNYTAMQE
                     QPGDAPRPTMSFIGNSASHPPQVSCWITQTTERTHEIIRAALHRSPLYSGQIEGAGPR
                     YCPSIEDKVVRFAEKNSHQIFVEPEGLNVIDIYPNGISTSLPFDVQLELVRSIRGFEQ
                     AHITRPGYAIEYDFFDPRGLKASLETKAIAGLFFAGQINGTTGYEEAAAQGLLAGLNA
                     ARHVRGLSSWTPRRDQAYLGVLVDDLITHGTNEPYRMFTSRAEYRLQLREDNADARLT
                     AIGRDLGLIDDARWAHFNAKQEAVARECERLSAFWATPGNALGREVKETLGVTLSRET
                     NIIDLMKRPELDYAALMRVPSLGPGVDDAQVAEQVEISVKYAGYLNRQSEEITRQQRH
                     EATAIPLEFDYAAVRGLSTEVLQKLQHIQPQTVGQAQRIPGMTPAAISLLLVHLERMR
                     RNRVA"
     gene            complement(4348. .4560)
                     /gene="XF2107"
     CDS             complement(4348. .4560)
                     /gene="XF2107"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84906.1"
                     /db_xref="GI:9107235"
                     /translation="MFILLMCVHTHPLIDELVEAWRAGPVIRSMYNAIKQDGSSGITE
                     RLPVRWFSWGRASKIDATAAAILASV"
     tRNA            4928. .5001
                     /gene="tRNA-Cys"
                     /note="found by tRNAscan"
                     /product="tRNA-Cys"
     gene            4928. .5001
                     /gene="tRNA-Cys"
     gene            5207. .5392
                     /gene="XF2108"
     CDS             5207. .5392
                     /gene="XF2108"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84907.1"
                     /db_xref="GI:9107236"
                     /translation="MQADGTTIIGLDISTKRWATAFLSGQNPLSIEDKSDFVLGAALR
                     NSAKNDTTQRLAPIKTA"
     gene            complement(5566. .5892)
                     /gene="XF2109"
     CDS             complement(5566. .5892)
                     /gene="XF2109"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84908.1"
                     /db_xref="GI:9107237"
                     /translation="MTGAPRVRSSFWKRFKQRLGRVKRCRCFLPWSWRRCVSCLSRVI
                     AVRVTVAASCVWVENKKPVPPAAAVERCRAAAISRDEAAYRMTWHGVYDWLLGKLQAA
                     SELHAA"
     gene            complement(6308. .6646)
                     /gene="XF2110"
     CDS             complement(6308. .6646)
                     /gene="XF2110"
                     /note="hypothetical protein; identified by sequence

  Query Match             64.0%;  Score 16;  DB 1;  Length 13856;
  Best Local Similarity   100.0%;  Pred. No. 96;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gatctgctgtttgaag 23
          ||||||||||||||||
Db    2652 GATCTGCTGTTTGAAG 2667
```

```
RESULT  40
AE004580
LOCUS       AE004580    15247 bp    DNA             BCT       30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 141 of 529 of the complete
            genome.
ACCESSION   AE004580 AE004091
VERSION     AE004580.1  GI:9947468
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 15247)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 15247)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .15247
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(122. .640)
                     /gene="hc"
                     /note="PA1512"
     CDS             complement(122. .640)
                     /gene="hc"
                     /codon_start=1
                     /transl_table=11
                     /product="secreted protein Hcp"
                     /protein_id="AAG04901.1"
                     /db_xref="GI:9947469"
                     /translation="MATPAYMSITGTKQGLITAGAFTEDSVGNTYQEGHEDQVMVQGF
                     NHEVIIPRDPQSGQPTGQRVHKPVVITKVFDKASPLLLAALTSGERLTKVEIQWYRTS
                     AAGTQEHYYTTVLEDAIIVDIKDYMHNCQDPGNAHFTHLEDVHFTYRKITWTHEVSGT
                     SGSDDWRSPVAG"
     gene            complement(1126. .2403)
                     /gene="PA1513"
     CDS             complement(1126. .2403)
                     /gene="PA1513"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04902.1"
                     /db_xref="GI:9947470"
                     /translation="MEAHLIEWLNLLVRWIHMIVGIAWIGASFYFVWLENNLNRSNPR
                     EGLSGDLWAIHGGGIYHLEKYKLAPPKMPENLHWFKWEAYSTWMSGVVLLTIVFYLNP
                     TLYLIAPGSDLAPAAAIAIGIGSLVCGWFVYTILCDSALGKTPALLGLVLFVLLVAAA
                     YGLTQVFSGRGAYLHVGAIIGTIMVGNVFRVIMPAQRALVKAIEENREPDPVLPAKGL
                     LRSRHNNYFTLPVLFIMISNHFPSTYGSQYNWLILAAIAVLAVLVRHYFNTRHESNRF
                     AWALPAGALGMIALAFVTTPNYQQASAPLASTPAAAPAQPATGGDAPAAPAATQASNA
                     GAASGADFDKVHHVIQERCTVCHSAKPTSQLFSTAPGGIMLDTPQQIQQLAPKIQAQA
                     VASQIMPLGNITQMTQEERNLIGSWIAKGAQIN"
     gene            complement(2452. .2961)
```

/gene="PA1514"
CDS complement(2452. .2961)
/gene="PA1514"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04903.1"
/db_xref="GI:9947471"
/translation="MRTLKIEPLTKEAFAPFGDVIETAGSDYFMINNGSTRRYHKLAT
VETAQPEDNAIISIFSAEKLEMPLRIRMLERHPLGSQAFIPLLGNPFLVVVAPLGDVP
VPGLVRAFLTNGRQGVNYHRGVWHHPVLTIEKRDDFLVVDRSGSGNNCDEHFFTEDEQ
LLLDPQSNQ"
gene complement(3016. .4014)
/gene="alc"
/note="PA1515"
CDS complement(3016. .4014)
/gene="alc"
/codon_start=1
/transl_table=11
/product="allantoicase"
/protein_id="AAG04904.1"
/db_xref="GI:9947472"
/translation="MNAEHSPFRHYLDLADARLGSQVVAVSDEWFAPASRMLQAGEPV
WKEGVFDDSGKWMDGWETRRKRFEGHDQAVIRLGVPGVLKGVDIDTRFFTGNHPPAAS
LDGCFCAEGDPDDSTSWSEVLAAVGLQGDSHHYHPIDDERPWTHLRLNIYPDGGIARL
RLYGVPYRDWSNQPPGTALDLAAAVNGGRALACSDQHFGRMGNLLNPGRAINMGDGWE
TGRRRTPGHDWVIVALGHPGSIEAAVVDTLHFKGNYPESCSIQAAFVEGGNEARIEAQ
SLFWRELLPAQKLEMHQEHRFERHLNALGPITHVRLNIFPDGGVSRLRLFGRPQLP"
gene complement(4032. .4547)
/gene="PA1516"
CDS complement(4032. .4547)
/gene="PA1516"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04905.1"
/db_xref="GI:9947473"
/translation="MSRFQTLTPASLSREAFIEAFADIYEHSPWVAEKAYDLGIDDSL
NDIEGLHQRMADILLSASREAQLALINAHPDLAGKAAIRGELTASSTSEQAGAGIHEC
SAEEFARFTELNDAYKARFGFPFIKAVKGSNRHQILAAFEERIQHSADEEFATALAEI
NKIALFRLQQL"
gene complement(4544. .5470)
/gene="PA1517"
CDS complement(4544. .5470)
/gene="PA1517"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04906.1"
/db_xref="GI:9947474"
/translation="MSADYPRDLIGYGNNPPHPHWPGDARIALSFVLNYEEGGERCVL
HGDKESEAFLSEMVAAQPLQGVRHMSMESLYEYGSRAGVWRLLKLFKRRNVPLTVFAV
AMAAQRNPEVIRAMVADGHEICSHGYRWIDYQYMDEAQEREHMLEAIRILTELTGQRP
VGWYTGRTGPNTRRLVMEEGGFLYDSDTYDDDLPYWDPASTAEKPHLVIPYTLDTNDM
RFTQVQGFNNGEQFFQYLKDAFDVLYEEGATAPKMLSIGLHCRLIGRPARMAALERFI
QYAQSHDKVWFARREDIARHWHREHPFQETEA"
gene 5843. .6223
/gene="PA1518"
CDS 5843. .6223
/gene="PA1518"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04907.1"
/db_xref="GI:9947475"
/translation="MSASGRPQPMGRLTTHVLDSAHGCPGHGIKIELYRVEGQQLELI
ATTLTNHDGRCDQPVLQGEDFRPGVYQLVFNAGDYYRARGVQLPEPAFLDQVVLRFGI
ASADDHYHVPLLISPYSYSTYRGS"
gene complement(6464. .7813)
/gene="PA1519"
CDS complement(6464. .7813)
/gene="PA1519"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG04908.1"
/db_xref="GI:9947476"
/translation="MESTKQEEQGIYAATPPATGLLERLFKLRQHGTTVRTELAAGLT
TFITMAYIIFVNPNIMADAGIDHGAAFVATCLAAAALGCFLMGLYANWPVGLAPGMGLN
AFFTYTVVGTMGYSWQIALGAVFISGVMFMLLTFSRVREWLLNSIPRSLRFAMGAGVG
LFLGLIGLKTAGIVVASPATLIKLGHLTSPGPLLAALCFLMIAVLEYRRVFGGILISI
LSVTLVGWALGLVQYGGVFSAPPSLAPTFLAMDIAGAFNVTMISVILAFLFVHMFDTA
GTLMGVAQRAHLVKEDGRIENLSKAMKADSASSVFGGMLGVPPVTSYVESAAGVAAGG
RTGLTAVVVGVLFVAAMFFAPLAGMIPAFATAGALIYVAMLMMGGMAHIDWDEHTETI
PAIVTVIMMPLTFSVADGIALGFITYVAMKVLTGKHREVSASLYALFVIFVAKFIFL"
gene 8188. .8967
/gene="PA1520"
CDS 8188. .8967
/gene="PA1520"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG04909.1"
/db_xref="GI:9947477"
/translation="MTDQLQHIRKPPRTGKAARSGTQDEIVYAHIFDAILEQRLAPGT
KLSEEALGEIFGVSRTIIRRALSRLAHEQVVLLRPNRGAVVASPSIDEARQILFARRT
VERAITELATDNATAERLAELREMVVQEQSSFARGDRGAGIRLSGEFHLKLAEMARNA
PLVSFQRSLVSQTSLIIAQYESGGRSHCSFDEHNEILDAIEKGDKERAVTLMMHHMEH
IDSKLNLESDNASGDLHAVFSHLLGGKKKPRRAKADSDSAA"
gene complement(9021. .10325)
/gene="PA1521"
CDS complement(9021. .10325)
/gene="PA1521"
/codon_start=1
/transl_table=11
/product="probable guanine deaminase"
/protein_id="AAG04910.1"
/db_xref="GI:9947478"
/translation="MTRNAKAYRAAILHSIADPAEVGVERSYEYFEDGLLLVEDGKVA
RLGDAETLLGEIGEVEVFEYRDALITPGFIDTHIHFPQTGMIASYGEQLLDWLNTYTF
PTERQFGDQAHADQVAEIFLQELLRNGTTTALVFGSVHRQSVESLFEAARRLDLRLIA
GKVMMDRNAPDYLTDTAESSYRDSKALIERWHGQGRLLYAVTPRFAPTSTAEQLDMAA
RLLREHPGVYLHTHLSENLKEIEWVKELFPERSGYLDVYDHHGLLGPRSVFAHGVHLC
DGECQRLAETGSAVAFCPTSNLFLGSGLFDLPKLERYKVKVGLGTDVGAGTSFSQLQS
LNEAYKVMQLQGARLDPFKSLYLATLGGARALELDDRIGSFATSNEADFVVLDYHATP
LLSYRLSQAGSLAERLFALTILGDDRTVKETFAAGRSVHRRD"
gene complement(10395. .11234)
/gene="PA1522"
CDS complement(10395. .11234)
/gene="PA1522"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04911.1"
/db_xref="GI:9947479"
/translation="MNWISALAELQRSGEPCVLVTIIEERGSTPRNAGSKMVVSAERL
YDTIGGGHLEYRAQAIAREMLAARTQDTRLERFSLGASLGQCCGGATVLLFEPMGQPQ
AHIAVFGAGHVGRALVPLLASLPCRVRWIDSREHEFPALLPDGVEKVVNDEVLDEVER
MPPGSHFIVMTHNHPLDLELTAAILARNDFAYFGLIGSKTKRVKFEHRLRERGVDAER
LQRMRCPMGLEQVKGKLPAEIAISIAGEVIATYNAAFGLERKQGPSSVTRLTPESRRA
HES"

Query Match 64.0%; Score 16; DB 1; Length 15247;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 1254 GTTGCTGGATCTGCTG 1269

RESULT 41
HS314G4
LOCUS HS314G4 31557 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid 314G4 from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains

```
            ESTs and CpG islands.
ACCESSION   Z69667
VERSION     Z69667.1  GI:1200312
KEYWORDS    16p13.3; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 31557)
  AUTHORS   Thomas,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail enquires regarding this sequence:
            humquery@sanger.ac.uk
COMMENT     IMPORTANT:
            This sequence is not the entire insert of clone 314G4. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
            The true right end of clone 314G4 is at 31557.
            314G4 is from a 280kb clone contig extending from the telomere of
            16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology
            Unit, Institute of Molecular Medicine, Oxford.
            314G4 came from the Los Alamos, flow sorted human Chromosome 16
            libraries constructed by Norman Doggett (unpublished).
FEATURES             Location/Qualifiers
     source          1. .31557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.3"
                     /clone="LA16-314G4"
                     /clone_lib="LA16"
     repeat_region   428. .716
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    914. .1691
                     /note="expressed region - matched by multiple ESTs"
     misc_feature    1692. .2316
                     /partial
                     /note="match: 5' EST R51457 clone 39198"
     misc_feature    2290. .2810
                     /partial
                     /note="match: 5' EST T87094 clone 115315"
     misc_feature    complement(3171. .3720)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   5136. .5489
                     /note="3 copies of 118 mer 88 % conserved"
     misc_feature    5908. .6086
                     /note="match: EST Z39463 clone c-1ag06"
     repeat_region   7536. .7567
                     /partial
                     /note="Alu repeat: matches 1. .32 of consensus"
     repeat_region   7568. .7856
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   9446. .9734
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    12398. .13777
                     /note="putative CpG island"
     repeat_region   13241. .13275
                     /note="7 copies of 5 mer 86 % conserved"
     repeat_region   14561. .14677
                     /partial
                     /note="Alu repeat: matches 254. .136 of consensus"
     repeat_region   14680. .14971
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   14974. .15106
                     /partial
                     /note="Alu repeat: matches 145. .1 of consensus"
     repeat_region   15127. .15231
                     /partial
                     /note="Alu repeat: matches 116. .1 of consensus"
     repeat_region   15248. .15535
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   15607. .15917
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   16005. .16114
                     /partial
                     /note="Alu repeat: matches 108. .1 of consensus"
     repeat_region   16123. .16283
                     /partial
                     /note="Alu repeat: matches 308. .150 of consensus"
     misc_feature    17661. .18594
                     /note="putative CpG island"
     repeat_region   18912. .19019
                     /note="3 copies of 36 mer 89 % conserved"
     repeat_region   19514. .19621
                     /note="MER22 element fragment"
     misc_feature    19922. .20499
                     /note="match: 5' EST H21176 clone 173392"
     repeat_region   20504. .20539
                     /note="4 copies of 9 mer 97 % conserved"
     repeat_region   21472. .21766
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    25040. .25378
                     /note="expressed region - matched by multiple ESTs"
     misc_feature    complement(25620. .25910)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   26707. .26817
                     /partial
                     /note="Alu repeat: matches 298. .184 of consensus"
     repeat_region   26818. .26963
                     /partial
                     /note="Alu repeat: matches 160. .1 of consensus"
     misc_feature    complement(27039. .27209)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   30474. .30611
                     /note="3 copies of 46 mer 94 % conserved"
     misc_feature    31087. .31220
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   31496. .31557
                     /partial
                     /note="Alu repeat: matches 1. .62 of consensus"
BASE COUNT     5713 a   9969 c   9918 g   5957 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 92;  Length 31557;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db  10104 GTTGCTGGATCTGCTG 10119


RESULT  42
SC8520X
LOCUS       SC8520X     41200 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome XIII cosmid 8520.
ACCESSION   Z49705 Z49700 Z71257
VERSION     Z49705.1  GI:825556
KEYWORDS    ALD2; aldehyde dehydrogenase; ATPase; CBF3B; centromere DNA-binding
            protein; coiled-coil; DNA mismatch repair; DnaJ; Mitochondrial
            carrier protein; MLH1; SMP2; transfer RNA-Met; transfer RNA-Val.
SOURCE      Saccharomyces cerevisiae.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 41200)
  AUTHORS   Hunt,S. and Bowman,S.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41200)
  AUTHORS   Barrell,B. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Saccharomyces cerevisiae chromosome XIII
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
```

```
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            All CDS over 100 codons have been analysed. CDS that are completely
            overlapped and those that are overlapped by more than 50%
            of their length by a larger CDS have been omitted from this
            analysis.
            Details of the omitted CDS are available on request. The more
            significant matches with motifs in the PROSITE database are
            also included but some of these may be fortuitous. The length in
            codons and the calculated codon adaptation index (CAI)
            is given for each CDS.
            Cosmid 8520 is overlapped at the start by cosmid 9375, EMBL id
            SC9375,
            accession no. Z47071 and at the end by by cosmid 8010.
FEATURES              Location/Qualifiers
     source           1. .41200
                      /organism="Saccharomyces cerevisiae"
                      /strain="AB972"
                      /db_xref="taxon:4932"
                      /chromosome="XIII"
                      /clone="cosmid 8520"
                      /map="13R"
     misc_feature     1. .365
                      /note="overlap with cosmid 9375"
     CDS              <1. .514
                      /note="YM8520.01, len: > 170, CAI: 0.25, overlaps an
                      extends YM9375.22, SW:YMAL_YEAST P28625, previously noted
                      hypothetical 41.6 KD protein in IMP1 3' region (RF1095),
                      hypothetical transmembrane protein"
                      /codon_start=2
                      /product="unknown"
                      /protein_id="CAA89788.1"
                      /db_xref="GI:825557"
                      /db_xref="SWISS-PROT:P28625"
                      /translation="VVGICSKNSIERNKKLGYDYLVPYDEGSIVENVKKLKQSVLEND
                      KFDMIFDSVGNHDFFPVIDQFLKPKAKNSFYVTIAGNNKADYKNISWRDFVSLSSILK
                      AINPFKKYNWRFGHPYPPNNFIEVGNEMIKKGTYKPPIDSVYEFDQYKEAIDRLMSNR
                      AKGKVVVKMK"
     CDS              756. .2183
                      /note="YM8520.02, unknown, len: 475, CAI: 0.12"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89789.1"
                      /db_xref="GI:825558"
                      /db_xref="SWISS-PROT:Q03790"
                      /translation="MADLQKQENSSRFTNVSVIAPESQGQHEQQKQQEQLEQQKQPTG
                      LLKGLNGFPSAPQPLFMEDPPSTVSGELNDNPAWFNNPRKRAIPNSIIKRSNGQSLSP
                      VRSDSADVPAFSNSNGFNNVTFGSKKDPRILKNVSPNDNNSANNNAHSSDLGTVVFDS
                      NEAPPKTSLADWQKEDGIFSSKTDNIEDPNLSSNITFDGKPTATPSPFRPLEKTSRIL
                      NFFDKNTKTTPNTASSEASAGSKEGASTNWDDHAIIIFGYPETIANSIILHFANFGEI
                      LEDFRVIKDFKKLNSKNMSKSPSLTAQKYPIYTGDGWVKLTYKSELSKSRALQENGII
                      MNGTLIGCVSYSPAALKQLASLKKSEEIINNKTSSQTSLSSKDLSNYRKTEGIFEKAK
                      AKAVTSKVRNAEFKVSKNSTSFKNPRRLEIKDGRSLFLRNRGKIHSGVLSSIESDLKK
                      REQASKSKKSWLNRLNNWLFGWNDL"
     CDS              complement(2320. .4503)
                      /note="YM8520.03c, unknown, len: 727, CAI: 0.12, some
                      similarity to WP:T04A8.16 CE01070 calpain thiol protease,
                      (27.1% identity in 155 aa overlap)"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89790.1"
                      /db_xref="GI:825559"
                      /db_xref="SWISS-PROT:Q03792"
                      /translation="MNDWHEFNAAIKSIYCNAEGDSSSIINRLVGLAMKSEDSTFIEA
                      VLVLKENVSKVDKQLRFLWLTSTINSRFYPPIPISEASPVSWNKTEYCAPGTEELQRR
                      YPGRAKLQNEEDYSGGIEQCRDVPDCSLVASLINLRSKNLNLPLIKQISSTKYHVNLS
                      FNGSNKRLVTVDISQIPTSVDGKQLSLKSNDISDKIGELALLLVSKGTYSTDGSNISI
                      DTYRLSGFLPEITQVNSYPFEKLWKFHKSNLCLMGAGTGNRSNDMIKPLVENHDYSII
                      DITYDSRLVKLRDPRNSALNVEISYEQYLKNFKQLYLNWNQEKLFKRSQVLHFRYDTS
                      RYNKFSIVADKPLFHLVNNSKVTETVWLLLESHLQDEGSQENRSVSFLNEAPECTICP
                      IEPPVECGGNHIGLQLVKLRLDAETERLLYCYSTTNNNFSIHSFSVVKEICFQRLKDT
                      KSLFAKVLFSFPYEIEGKASFDTCNFFQNPTFELEVHSEQDYQVLMDAACISTSSHDL
                      INIQVYYFNDYELIKPIMFDNHYQPGQGLKQDVPILTNVKYMIVCSTYGPPASTEFEL
                      LASIRLSSSWRLISGITLRSVNLIYGTYPYHCRNRFHWKETSDKLKIQMTLPTKKYST
                      NKLFIRVVPVESSARLRMRCNIFEPESALCVYECQEYRTCPSGGIVIPDLEVSRTNIV
                      VLMIERSVPISSCLPTEGQLDELELFVGSSQKIRIEKYSDDVIPK"
     CDS              4872. .6515
                      /note="YM8520.04, unknown, len: 547, CAI: 0.10, potential
                      integral membrane protein"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89791.1"
                      /db_xref="GI:825560"
                      /db_xref="SWISS-PROT:Q03795"
                      /translation="MVKKHQNSKMGNTNHFGHLKSFVGGNVVALGAGTPYLFSFYAPQ
                      LLSKCHIPVSASSKLSFSLTIGSSLMGILAGIVVDRSPKLSCLIGSMCVFIAYLILNL
                      CYKHEWSSTFLISLSLVLIGYGSVSGFYASVKCANTNFPQHRGTAGAFPVSLYGLSGM
                      VFSYLCSKLFGENIEHVFIFLMVACGCMILVGYFSLDIFSNAEGDDASIKEWELQKSR
                      ETDDNIVPLYENSNDYIGSPVRSSSPATYETYALSDNFQETSEFFALEDRQLSNRPLL
                      SPSSPHTKYDFEDENTSKNTVGENSAQKSMRLHVFQSLKSSTFIGYYIVLGILQGVGL
                      MYIYSVGFMVQAQVSTPPLNQLPINAEKIQSLQVTLLSLLSFCGRLSSGPISDFLVKK
                      FKAQRLWNIVIASLLVFLASNKISHDFSSIEDPSLRASKSFKNISVCSAIFGYSFGVL
                      FGTFPSIVADRFGTNGYSTLWGVLTTGGVFSVSVFTDILGRDFKANTGDDDGNCKKGV
                      LCYSYTFMVTKYCAAFNLLFVLGIIGYTYYRRRATANSL"
     CDS              complement(6621. .7337)
                      /note="YM8520.05c, unknown, len: 238, CAI: 0.15"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89792.1"
                      /db_xref="GI:825561"
                      /db_xref="SWISS-PROT:Q03796"
                      /translation="MSHKLTILPFLIKFTPKFPQSIDHDEHGLNVYAFDLDHTIIKPK
                      SPNISFSRSASDWQFINFNSKKSTLDYLCNIIDNDPTAVIVIFSNQGGVITVPRTSKS
                      CTKYTNKILLFLKAIKNDERGETLSHRLWLYAAPKRPKTFAANHSKITFASLGESYNN
                      DPNIFEKVRKPMTGMVEFFKRDLESAYRVSEQISPIKLNWIYYCGDAAGRKKDFSDSD
                      IKFAENLHVEFKYPEEIFHG"
     CDS              complement(7598. .8365)
                      /note="YM8520.06c, unknown, len: 255, CAI: 0.17"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89793.1"
                      /db_xref="GI:825562"
                      /db_xref="SWISS-PROT:Q03798"
                      /translation="MLRPLRKSVLASCRHCFKVCGGLPQKQLPLFSPLLLRARYSSTD
                      SSTKRSNKSDKIDAPGFKKIFLVAIIGTVIFVKTVQSLDKNKPKTTLSEEEFENVVKG
                      LKRRVAIFPQGEVDIKFSLSPSIEETRKVLQKSQGDDINELQFVDPVKVIDYYRTLRD
                      DRYEALLNEYYKKYGCDTYAYNLPTGMLVMLLGRYFKENFKTGDKLVVVNFPHSIADA
                      TRFENEVSIVSKIFVPRKLSGSDVCKYYETVGKADII"
     CDS              8569. .9036
                      /note="YM8520.07, unknown, len; 155, CAI: 0.13"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89794.1"
                      /db_xref="GI:825563"
                      /db_xref="SWISS-PROT:Q03799"
                      /translation="MSLVKLANTCAHLQNCSKVRVALTSIPYTKLQLQFAYNLYQQGF
                      LSSLQKGSTMGPDKDFVEVTPDNISTRRLWVGLKYRDNKPVLSSCKLISKPNSRIHLP
                      MEDMKKLCSGVTIRNIKPLQPGELILVRAHNNIMDINEAISKKLDGEVLCRVK"
     tRNA             complement(9204. .9276)
                      /gene="tRNA-Met"
                      /note="tRNA-Met gene, anticodon cat, len: 73"
     gene             complement(9204. .9276)
                      /gene="tRNA-Met"
     LTR              complement(9432. .9659)
                      /note="remnant of solo delta element"
     CDS              complement(10797. .11249)
                      /note="YM8520.08c, unknown, len: 150, CAI: 0.13, possible
                      coiled-coil protein"
                      /codon_start=1
                      /product="unknown"
                      /protein_id="CAA89795.1"
                      /db_xref="GI:825564"
                      /db_xref="SWISS-PROT:Q03818"
                      /translation="MGNFIITERKKAKEERSNPQTDSMDDLLIRRLTDRNDKEAHLNE
                      LFQDNSGAIGGNIVSHDDALLNTLAILQKELKSKEQEIRRLKEVIALKNKNTERLNDE
                      LISGTIENNVLQQKLSDLKKEHSQLVARWLKKTEKETEAMNSEIDGTK"
```

```
     CDS             11387. .13837
                     /note="YM8520.09, unknown, len: 816, CAI: 0.13"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA89796.1"
                     /db_xref="GI:825565"
                     /db_xref="SWISS-PROT:Q03823"
                     /translation="MDVVQQKQDKQLQHQTQEQQQIREDQQEVPPQRPRQQNRWKPWW
                     NSTADDEPNTGRIAEYPNGQGRSSPTTDFQDSVNSNNDNKGIWSKIASFATSRYRSAP
                     IVVDDNTRYSQLNTEQINFLENEAKDIISKKSKSWCWYEAIPHISNSSNIIDSIDTPG
                     IISVSGTGSAKCPLPLNKYPGEGGNPGYNVFINDSLILPSDNPLNFLHVQPLRTKVLN
                     TIKNYYNFPNEQHLYLRQKKTALLKDKRIIIISVVGDLPEKYEQRSLESQRSAYYLSR
                     KLSQNLAQEQPQRVLTLSFQCPLHNQDLIPTYKECVELLNHWAHLFKEVDSIFFVGVY
                     HSVPLTLLLAKYIVQNNEVLEFDENTTVSVLSFQSCLQGYRFWDHSTDFTSNSYNNLG
                     SNSSTNENDSNDHDSNNDFTTKSQQIKEKQLFQGIDKKQQDTLSKIKNYRRIDSSESK
                     LVQDALDWLLFNWDTFRLTFFGKLYDNFMTISEKLAIDYNHPKILRNLWCNGKYMGID
                     LKNANNLNLDTDDEATSNINDVHVRTPNFESRLKIPTNRLFEITLWDILMITENLGYK
                     QFIPIINLLSPFFISRSFNDYTLPPNIRKQYQNSNKIWLQEMDSKWKMNGHQLNYDQR
                     EGESLGSSSESLLPENISTVKDFLQFVQYQNEKSSDFVRIYSDIYDDDKVYKCFLYNT
                     IFTKNPLSRKHLRLNIDLDTPTSILNTVNQYDLVWKIHDSFSKLIQLKNLPQREIPHA
                     LRLSISLNCFLDSTTSTSGPVFQRDTVEALRRLTEIWRTYQDWSPPTRGLKHLRDILS
                     VLAMYDNPKNLINDVRRT"
     CDS             14039. .14713
                     /note="YM8520.10, putative DNAJ gene, len: 224, CAI: 0.14,
                     similar to DNAJ proteins"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA89797.1"
                     /db_xref="GI:825566"
                     /db_xref="SWISS-PROT:P48353"
                     /translation="MSFTEDQEKIALEILSKDKHEFYEILKVDRKATDSEIKKAYRKL
                     AIKLHPDKNSHPKAGEAFKVINRAFEVLSNEEKRSIYDRIGRDPDDRQMPSRGAASGF
                     RGSAGGSPMGGGFEDMFFNSRFGGQRAGPPEDIFDFLFNAGGSPFGASPFGPSASTFS
                     FGGPGGFRVYTNNRGGSPFMRQQPRSRQQQQQAEENAVNSQLKNMLVLFIIFIVLPMI
                     KDYLFS"

  Query Match             64.0%;  Score 16;  DB 14;  Length 41200;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgtt 18
           ||||||||||||||||
Db   25021 TGCTGGATCTGCTGTT 25036


RESULT  43
AX008043/c
LOCUS       AX008043    46340 bp    DNA             PAT       06-SEP-2000
DEFINITION  Sequence 66 from Patent WO9967395.
ACCESSION   AX008043
VERSION     AX008043.1  GI:9995728
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46340)
  AUTHORS   Perin,J.P., Rieger,F. and Alliel,P.M.
  TITLE     Nucleic sequence and deduced protein sequence family with human
            endogenous retroviral motifs, and their uses
  JOURNAL   Patent: WO 9967395-A 66 29-DEC-1999;
            INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
            FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES             Location/Qualifiers
     source          1. .46340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    16104 a   8737 c   8435 g  13064 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 46340;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgttt 19
           ||||||||||||||||
Db   1254 GCTGGATCTGCTGTTT 1239


RESULT  44
AB018119
LOCUS       AB018119    62927 bp    DNA             PLN       27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MSN2.
ACCESSION   AB018119 BA000015
VERSION     AB018119.1  GI:3702737
KEYWORDS    .
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MSN2.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (1), 31-63 (2000)
  MEDLINE   20181125
REFERENCE   2  (bases 1 to 62927)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MSN2
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is K1F13 and the 3' clone is MUD21.
FEATURES             Location/Qualifiers
     source          1. .62927
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MSN2"
                     /clone_lib="Mitsui P1"
     CDS             join(5223. .5458,5553. .5649,5719. .5799,5902. .6021,
                     6101. .6313,6392. .6454,6544. .6721,6803. .7170)
                     /note="emb|CAB16816.1
                     gene_id:MSN2.2
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA97270.1"
                     /db_xref="GI:8843722"
                     /translation="MVRRKRQEEDEKIEIERVKEESLKLAKQAEEKRRLEESKEQGKR
```

```
                IQVDDDQLAKTTSKDKGQINHSKDVVEEDVNPPPSSIDGKSEIGDGTSVNPRCLCCFH
                CHRPFVMHEILKKGKFHIDCYKEYYRNRNCYVCQQKIPVNAEGIRKFSEHPFWKEKYC
                PIHDEDGTAKCCSCERLEPRGTNYVMLGDFRWLCIECMGSAVMDTNEVQPLHFEIREF
                FEGLFLKVDKEFALLLVEKQALNKAEEEEKIDYHRAAVTRGLCMSEEQIVPSIIKGPR
                MGPDNQLITDIVTESQRVSGFEVTGILIIYGLPRLLTGYILAHEMMHAWLRLNGYKNL
                KLELEEGLCQALGLRWLESQTFASTDAAAAAAVASSSSFSSSTAPPAAITSKKSDDWS
                IFEKKLVEFCMNQIKEDDSPVYGLGFKQVYEMMVSNNYNIKDTLKDIVSASNATPDST
                V"
CDS             complement(join(8609. .9043,9132. .9662))
                /note="contains similarity to unknown protein
                gb|AAC17084.1
                gene_id:MSN2.3"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97271.1"
                /db_xref="GI:8843723"
                /translation="MSSKKSLVQSLFNISKTYSRISGLTRMRPTKSGGIPPDAGDSGI
                RRRFLHKRAFFSPEIVPKGGNLMEKLRELTLSNNNRIRLDEMLPPPSPKKSSPEFFPA
                VTVEDVKKLMRAAEMELVKSKLREIGKNWVPYSEFVRVCGEYSSDPEQGNRVANMLDE
                AGNVIVLGKLVCLKPEELTSAMAGLIPTLEPSLDAETRQEFEQLEIIKSDIDKRADDL
                VRKELWAGLGLIMAQTVGFFRLTFWELSWDVMEPICFYVTSTYFMAGYAFFLRTSKEP
                SFEGFYKSRFETKQKRLIKMLDFDIDRFTKLQKMHRPNLTKSGRC"
CDS             complement(14713. .15909)
                /note="gene_id:MSN2.4"
                /codon_start=1
                /evidence=not_experimental
                /product="At14a protein-like"
                /protein_id="BAA97272.1"
                /db_xref="GI:8843724"
                /translation="MEFCGCLSELMNGESSSKRNGPSTLPVKEVRTDMRSKYSSDLSS
                YTSACKKDSNLKSFDSSLHQRTNIIITSLAARAETQSLNLDSLMEVYGFLLELNQNAV
                RVIIESREDVWKNKDLKSLVDVYFKSTSKTLDFCNTVENCVKRTEISQLIIRFAVKQF
                EAESVDTDLGGDKKKKKYTKTLEELNKFKAMGDPFDGELVTQFDSVYDQQVLFLEELR
                KQRRKLDKKQRNVKTLRTVSNVFFATAYVSVLVLSVVATTMSAPPVVCAVASGSTAPI
                EITGKWFSQMWKKYEKAVKRQRGLVLTMESRVQVNNEAMKNIRSDVDELRSWVSSILE
                TVDFAVEREEEEEEAMGLAMQGIKKHVDGFTEKMEEVGENAAKCSKFIALGRLLVLEHI
                LGLPAN"
CDS             complement(16879. .18105)
                /note="gene_id:MSN2.5"
                /codon_start=1
                /evidence=not_experimental
                /product="At14a protein-like"
                /protein_id="BAA97273.1"
                /db_xref="GI:8843725"
                /translation="MVFCGLFSELIKGHSSSNNNGTNVSLAKTIQTNIRSQYSSDLSS
                YASACKKYSSLKSFDSLLHERTNSIISSLAAQAKTRSLNIESLMEVYGYLLELNQDTV
                RVIIESKEDVLKNNDLKALVDVYFKSTSKTLDFCNTVEKCVKKAEISQLIIRFAVKQF
                ETETVDTDLGESKKKKYVKTLEEMNKFKAMGDPFDGEFVTQYKSVYDEQVLLLDELRK
                LKVKLGKKLRNIKTWRILSNVVFATAFVTVFVLSVVAAAMMAPPVLSAVASGLTTPIE
                VVGMWCNKMWKEYEKAVKRQRGLVLTMELGVQANNVTMVNIKFEVENLSIRISSILKT
                VNFAVDREENEMATRFAMQEIKKKVEGFTEKIEEVGERAANCSKLIALGRLVVLGHIL
                GLHIVEGGAANIISSV"
CDS             complement(20757. .21976)
                /note="gene_id:MSN2.6"
                /codon_start=1
                /pseudo
                /evidence=not_experimental
                /product="At14a protein-like"
CDS             complement(join(23245. .23292,23403. .23486,23567. .23671,
                23788. .23913,24012. .24083,24250. .24325,24468. .24509,
                24594. .24649,24801. .24929,25065. .25118,25274. .25399,
                25486. .25622,25713. .25986))
                /note="gene_id:MSN2.7"
                /codon_start=1
                /evidence=not_experimental
                /product="dolichyl-di-phosphooligosaccharide-protein
                glycotransferase (oligosaccharyltransferase)-like"
                /protein_id="BAA97274.1"
                /db_xref="GI:8843726"
                /translation="MVNLSRSVALISVFLLPLLSFSFSVDNPTDRRVLVLLDDLSLKS
                SHSIFFNTLKSRGFDLDFKLAEDSKLALQRYGQYLYDGLIIFAPSTERFGGSLDSKSI
                ADFVDSGKDLILSADTAASDLIRGIATECGVDFDEDSSAMVIDHTSFSVSDVDGDHTL
                IAADDLVKSDVILGKTKIEAPVLFRGVAHSLNPTNNLVLKVLSASPSAYSANPSSKLS
                SPPQLTGSSISLVSVMQVCLSSARNNARVVISGSVQLFSDRLIRSGVQKAGSPNQYEK
                SGNEQFVTELSKWVFHERGHLKAGNLVHHRVGETDEPAIYRIKDDLEFSVEIYEWSGK
                SWEPYVANDVQVQFYMMSPYVLKTLSTDKKGLFHTSFKVPDVYGVFQFKVEYEKLGYT
                TLSLSKQIPVRPYRHNEYERFIPTAYPYYGACFTTMAGFFVFSFVYLYHK"
CDS             30560. .32005
                /note="gene_id:MSN2.8"
                /codon_start=1
                /evidence=not_experimental
                /product="UTP-glucose glucosyltransferase"
                /protein_id="BAA97275.1"
                /db_xref="GI:8843727"
                /translation="MHITKPHAAMFSSPGMGHVIPVIELGKRLSANNGFHVTVFVLET
                DAASAQSKFLNSTGVDIVKLPSPDIYGLVDPDDHVVTKIGVIMRAAVPALRSKIAAMH
                QKPTALIVDLFGTDALCLAKEFNMLSYVFIPTNARFLGVSIYYPNLDKDIKEEHTVQR
                NPLAIPGCEPVRFEDTLDAYLVPDEPVYRDFVRHGLAYPKADGILVNTWEEMEPKSLK
                SLLNPKLLGRVARVPVYPIGPLCRPIQSSETDHPVLDWLNEQPNESVLYISFGSGGCL
                SAKQLTELAWGLEQSQQRFVWVVRPPVDGSCCSEYVSANGGGTEDNTPEYLPEGFVSR
                TSDRGFVVPSWAPQAEILSHRAVGGFLTHCGWSSTLESVVGGVPMIAWPLFAEQNMNA
                ALLSDELGIAVRLDDPKEDISRWKIEALVRKVMTEKEGEAMRRKVKKLRDSAEMSLSI
                DGGGLAHESLCRVTKECQRFLERVVDLSRGA"
CDS             join(39811. .40266,40937. .41167)
                /note="gene_id:MSN2.9"
                /codon_start=1
                /evidence=not_experimental
                /product="homeodomain transcription factor-like"
                /protein_id="BAA97276.1"
                /db_xref="GI:8843728"
                /translation="MDHGRLMDDQMMLGSQVYPYTTQPQNSHCIIVNQIDGGEESKPV
                KRRRKRRSKGSSATNEEDVAEIGGMLRKRKLTDEQVNMLEYSFGNEHKLESGRKEKIA
                GELGLDPRQVAVWFQNRRARWKNKKLEEEYAKLKNHHDNVVLGQCQLESQILKLTEQL
                SEAQSEIRKLSERLEEMPTNSSSSSLSVEANNAPTDFELAPETNYNIPFYMLDNNYLQ
                SMEYWDGLYV"
CDS             join(42014. .42279,42692. .42827,42933. .43128,43201. .43322,
                43403. .43558,43628. .43969)
                /note="gene_id:MSN2.10"
                /codon_start=1
                /evidence=not_experimental
                /product="protein kinase ATN1-like protein"
                /protein_id="BAA97277.1"
                /db_xref="GI:8843729"
                /translation="MRPRGYQRAPSMQKPTDYPTDKTLHPNYPFLMSSHGLKSFESDD
                EDDDSDSSNDQFAFTINTELLVDVKDISIGDFIGEGSSSTVYRGLFRRVVPVSVKIFQ
                PKRTSALSIEQRKKFQREVLLLSKFRHENIVRFIGACIEPKLMIITELMEGNTLQKFM
                LSVRPKPLDLKLSISFALDIARGMEFLNANGIIHRDLKPSNMLLTGDQKHVKLADFGL
                AREETKGFMTFEAGTYRWMAPELFSYDTLEIGEKKHYDHKVDVYSFAIVFWELLTNKT
                PFKGKNNIFVAYAASKNQRPSVENLPEGVVSILQSCWAENPDARPEFKEITYSLTNLL
                RSLSSDTDATSSNSKANIATEDSTSSLVQERVVCDCPGLKMSKTKKLKKKTNKLMNMI
                VPFLKIFKSCMSK"
CDS             complement(join(44420. .44707,44833. .44982,45075. .45384,
                45454. .45950))
                /note="contains similarity to unknown protein
                emb|CAB46038.1
                gene_id:MSN2.11"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97278.1"
                /db_xref="GI:8843730"
                /translation="MSATALSRLNPVSQFGFQRIVAGKSKSFFSNSGQRRLFSDSSRF
                RQAMAASGSLPVFGDACLDDLVTTCSNGLDFTKKRSSGGSFTINCPVASMRLGKRGGM
                MKNRLVCHYSVVDPLEKSRALFGTLSKSVHTSPMACFSVGPAHELSSLNGGSQESPPT
                TTTSLKSLRLVSGSCYLPHPEKEATGGEDAHFICDEEQAIGVADGVGGWAEVGVNAGL
                FSRELMSYSVSAIQEQHKGSSIDPLVVLEKAHSQTKAKGSSTACIIVLKDKGLHAINL
                GDSGFTVVREGTTVFQSPVQQHGFNFTYQLESGNSADVPSSGQVFTIDVQSGDVIVAG
```

```
  Query Match             64.0%;  Score 16;  DB 12;  Length 62927;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgtt 18
           ||||||||||||||||
Db   55052 TGCTGGATCTGCTGTT 55067
```

```
RESULT  45
AC013956/c
LOCUS       AC013956    65679 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013956
VERSION     AC013956.1  GI:6437379
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 65679)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211528 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .65679
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    18648 a  14529 c  14360 g  18142 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 63;  Length 65679;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db   63993 GTTGCTGGATCTGCTG 63978



Search completed: October  9, 2001, 15:44:56
Job time: 14682 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:06 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 41608

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	20	AAX01303	E. coli biotin syn
	2	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	3	25	100.0	1121	7	AAN60496	Sequence encoding
	4	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	5	16	64.0	5000	21	AAZ36139	Nucleotide sequenc
c	6	15	60.0	1683	16	AAT14850	Maize U14.4 DNA se
c	7	15	60.0	1683	16	AAT08746	Maize U14.4 DNA se
c	8	15	60.0	1689	20	AAZ23024	Rat kd312 polypept
c	9	15	60.0	2022	21	AAC44132	Zea mays DNA fragm
c	10	15	60.0	2034	20	AAX04606	Maize genomic clon
c	11	15	60.0	3079	20	AAZ23025	Rat kd312 genomic
	12	15	60.0	6703	19	AAV49536	Adenylate cyclase
	13	14	56.0	275	20	AAZ12718	Human gene express
c	14	14	56.0	336	22	AAF65487	Novel human polynu
c	15	14	56.0	379	20	AAV87319	EST clone BP797.
c	16	14	56.0	383	21	AAC26529	Human secreted pro
c	17	14	56.0	384	18	AAV01888	Human OTK27 gene.
	18	14	56.0	612	21	AAC99862	Human secreted pro
c	19	14	56.0	619	21	AAA44124	Human secreted exp
c	20	14	56.0	707	21	AAF14018	Aspergillus oryzae
	21	14	56.0	735	16	AAQ92502	Mouse antibody F4-
c	22	14	56.0	753	20	AAZ00347	Nucleotide sequenc
c	23	14	56.0	769	21	AAF15012	Trichoderma reesei
	24	14	56.0	891	20	AAZ06233	Human secreted pro
c	25	14	56.0	1044	21	AAX82834	L-asparaginase DNA
	26	14	56.0	1155	20	AAZ09784	E. coli S-adenosyl
	27	14	56.0	1155	22	AAA99433	Met K oligonucleot
c	28	14	56.0	1173	20	AAZ00358	Nucleotide sequenc
	29	14	56.0	1220	21	AAC44861	Arabidopsis thalia
c	30	14	56.0	1367	21	AAA81673	N. meningitidis pa
c	31	14	56.0	1493	18	AAV01889	Human OTK27 gene.
c	32	14	56.0	1676	22	AAD03029	Flax 2S storage pr
c	33	14	56.0	1848	18	AAT96346	Chimeric gene cont
c	34	14	56.0	1998	20	AAZ00367	Nucleotide sequenc
	35	14	56.0	2080	20	AAZ30162	cDNA encoding a po
c	36	14	56.0	2087	20	AAZ00343	Nucleotide sequenc
c	37	14	56.0	2166	20	AAZ00379	Nucleotide sequenc
c	38	14	56.0	2405	22	AAF27417	Human fatty acid t
c	39	14	56.0	2574	20	AAZ33977	Human PRO703 nucle
c	40	14	56.0	2574	21	AAC78481	Human PRO703 (UNQ3
c	41	14	56.0	2574	21	AAC58239	Human PRO703 nucle
c	42	14	56.0	2669	20	AAZ25020	Human H-beta 58 fa
c	43	14	56.0	2675	21	AAC98021	Human colon cancer
c	44	14	56.0	3243	11	AAQ05686	Islets of Langerha
c	45	14	56.0	3243	20	AAZ32338	Human pancreatic i

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
DR P-PSDB; AAW73906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     57 gttgctggatctgctgtttgaagcg 81


RESULT   2
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     80 gttgctggatctgctgtttgaagcg 104


RESULT   3
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     98 gttgctggatctgctgtttgaagcg 122


RESULT   4
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
```

```
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 gttgctggatctgctgtttgaagcg 197


RESULT   5
AAZ36139
ID   AAZ36139 standard; DNA; 5000 BP.
XX
AC   AAZ36139;
XX
DT   11-FEB-2000  (first entry)
XX
DE   Nucleotide sequence of the DDM1 gene which regulates DNA methylation.
XX
KW   DDM1 gene; chromosome 5; zinc-finger protein gene; SWI2/SNF2 family;
KW   glutamic acid tRNA gene; genomic cytosine methylation; gene silencing;
KW   DNA methylation; chromatin structure; DNA hypomethylation;
KW   DDM1 gene allele; inbreeding depression; ss.
XX
```

```
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             535..4354
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            535..566
FT                   /*tag= b
FT                   /number= 1
FT   intron          567..771
FT                   /*tag= c
FT                   /number= 1
FT   exon            772..850
FT                   /*tag= d
FT                   /number= 2
FT   allele          replace(785, (N82))
FT                   /*tag= e
FT                   /note= "site of the ddm1-5 (som8) mutation; this base is
FT                          replaced with 82 unspecified bases"
FT   intron          851..985
FT                   /*tag= f
FT                   /number= 2
FT   exon            986..1252
FT                   /*tag= g
FT                   /number= 3
FT   intron          1253..1353
FT                   /*tag= h
FT                   /number= 3
FT   exon            1354..1440
FT                   /*tag= i
FT                   /number= 4
FT   intron          1441..1548
FT                   /*tag= j
FT                   /number= 4
FT   exon            1549..1895
FT                   /*tag= k
FT                   /number= 5
FT   intron          1896..1975
FT                   /*tag= l
FT                   /number= 5
FT   exon            1976..2165
FT                   /*tag= m
FT                   /number= 6
FT   intron          2164..2250
FT                   /*tag= n
FT                   /number= 6
FT   exon            2251..2426
FT                   /*tag= o
FT                   /number= 7
FT   allele          replace(2384..2385)
FT                   /*tag= p
FT                   /note= "site of the ddm1-6 mutation; either the G
FT                          at 2384 or 2385 is deleted"
FT   intron          2427..2558
FT                   /*tag= q
FT                   /number= 7
FT   exon            2559..2625
FT                   /*tag= r
FT                   /number= 8
FT   intron          2626..2702
FT                   /*tag= s
FT                   /number= 8
FT   exon            2703..2892
FT                   /*tag= t
FT                   /number= 9
FT   intron          2893..2974
FT                   /*tag= u
FT                   /number= 9
FT   exon            2975..3070
FT                   /*tag= v
FT                   /number= 10
FT   intron          3071..3147
FT                   /*tag= w
FT                   /number= 10
FT   exon            3148..3242
FT                   /*tag= x
FT                   /number= 11
FT   misc_feature    3186
FT                   /*tag= y
FT                   /note= "alternative splice donor site used in ddm1-2"
FT   intron          3243..3316
FT                   /*tag= z
FT                   /number= 11
FT   allele          replace(3243, A)
FT                   /*tag= aa
FT                   /note= "site of the ddm1-2 mutation"
FT   exon            3317..3436
FT                   /*tag= ab
FT                   /number= 12
FT   allele          replace(3337, A)
FT                   /*tag= ac
FT                   /note= "site of the ddm1-7 (som5) mutation"
FT   intron          3437..3539
FT                   /*tag= ad
FT                   /number= 12
FT   exon            3540..3659
FT                   /*tag= ae
FT                   /number= 13
FT   intron          3660..3744
FT                   /*tag= af
FT                   /number= 13
FT   exon            3745..3843
FT                   /*tag= ag
FT                   /number= 14
FT   intron          3844..3933
FT                   /*tag= ah
FT                   /number= 14
FT   exon            3934..4038
FT                   /*tag= ai
FT                   /number= 15
FT   intron          4039..4129
FT                   /*tag= aj
FT                   /number= 15
FT   exon            4130..4354
FT                   /*tag= ak
FT                   /number= 16
FT   tRNA            4755..4826
FT                   /*tag= al
FT                   /note= "complement of predicted tRNA-glu"
XX
PN   WO9955891-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09268.
XX
PR   24-JUN-1998;   98US-0104070.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Richards EJ,  Jeddeloh JA;
XX
DR   WPI; 2000-013439/01.
DR   P-PSDB; AAY43847.
XX
PT   Arabidopsis thaliana DDM1 gene alters methylation state of DNA, useful
PT   for reducing gene silencing or inbreeding depression
XX
PS   Claim 6; Page 36-38; 59pp; English.
XX
CC   The present sequence represents the DDM1 gene of Arabidopsis thaliana.
CC   The gene is located on chromosome 5, lower arm, where it is flanked on
CC   the centromeric side, within 20 kb, by a zinc-finger protein gene and
CC   on the telomeric side, within 1 kb, by a glutamic acid tRNA gene. The
CC   DDM1 protein is a member of the SWI2/SNF2 family of polypeptides. The
CC   DDM1 protein is important for the maintenance of proper genomic cytosine
```

```
CC   methylation, and its function appears to be necessary to maintain gene
CC   silencing. The protein appears to function in the DNA methylation system
CC   by affecting chromatin structure. Disruption of the gene is associated
CC   with DNA hypomethylation. Four DDM1 gene alleles have been identified,
CC   all of which are expected to destroy or severely reduce gene function.
CC   The DDM1 nucleic acid can be used to create transgenic plants with
CC   altered levels of DNA methylation. The mutant or engineered plants are
CC   expected to be viable and display valuable characteristic such as
CC   inbreeding depression or a reduction in gene silencing.
XX
SQ   Sequence 5000 BP; 1386 A; 909 C; 1029 G; 1676 T; 0 other;


  Query Match           64.0%;  Score 16;  DB 21;  Length 5000;
  Best Local Similarity  100.0%;  Pred. No. 7.4;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgtt 18
          ||||||||||||||||
Db   1132 tgctggatctgctgtt 1147


RESULT   6
AAT14850/c
ID   AAT14850 standard; DNA; 1683 BP.
XX
AC   AAT14850;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Maize U14.4 DNA sequence.
XX
KW   snoRNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW   Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW   normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
KW   PCR; polymerase chain reaction; amplification; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    379..385
FT                   /*tag=  a
FT                   /note= "nucleotides not identified in sequence"
FT   repeat_unit     1212..1217
FT                   /*tag=  b
FT                   /note= "inverted repeat"
FT   misc_RNA        1218..1329
FT                   /*tag=  c
FT                   /note= "MU14.4"
FT   misc_feature    1220..1225
FT                   /*tag=  d
FT                   /note= "Box C"
FT   misc_feature    1324..1329
FT                   /*tag=  e
FT                   /note= "Box D"
FT   repeat_unit     1330..1335
FT                   /*tag=  f
FT                   /note= "inverted repeat"
XX
PN   WO9530748-A2.
XX
PD   16-NOV-1995.
XX
PF   04-MAY-1995;   95WO-EP01694.
XX
PR   09-DEC-1994;   94EP-0119487.
PR   04-MAY-1994;   94WO-EP01409.
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ,  Waugh R;
XX
DR   WPI; 1995-404115/51.
XX
PT   Plant U14 nucleic acid sequences and derivatives - regulate
PT   ribosomal RNA production and accumulation in plants
XX
PS   Example; Fig 7A-D; 57pp; English.
XX
CC   The present sequence shows the maize single U14 sequence U14.4,
CC   including flanking intergenic sequences and inverted repeats, as obtd.
CC   by sequencing of fragments of isolated genomic clones. Inverted repeats
CC   and Box C and D elements are shown above (see AAT08751-52). U14 genes
CC   are small nucleolar RNAs present in the genome as non-translated
CC   sequences. SnoRNAs are involved in processing of pre-ribosomal RNA
CC   transcripts and ribosome formation. U14snoRNA is required for a normal
CC   growth phenotype and for processing of pre-rRNA transcripts in yeast.
CC   Plant U14 sequences have been identified and characterised in the present
CC   invention, and may be intron encoded.
XX
SQ   Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;


  Query Match           60.0%;  Score 15;  DB 16;  Length 1683;
  Best Local Similarity  100.0%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgct 15
          |||||||||||||||
Db    601 GTTGCTGGATCTGCT 587


RESULT   7
AAT08746/c
ID   AAT08746 standard; DNA; 1683 BP.
XX
AC   AAT08746;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Maize U14.4 DNA sequence.
XX
KW   snoRNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW   Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW   normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
KW   PCR; polymerase chain reaction; amplification; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    379..385
FT                   /*tag=  a
FT                   /note= "nucleotides not identified in sequence"
FT   repeat_unit     1212..1217
FT                   /*tag=  b
FT                   /note= "inverted repeat"
FT   misc_RNA        1218..1329
FT                   /*tag=  c
FT                   /note= "MU14.4"
FT   misc_feature    1220..1225
FT                   /*tag=  d
FT                   /note= "Box C"
FT   misc_feature    1324..1329
FT                   /*tag=  e
FT                   /note= "Box D"
FT   repeat_unit     1330..1335
FT                   /*tag=  f
FT                   /note= "inverted repeat"
XX
PN   WO9530747-A1.
XX
PD   16-NOV-1995.
XX
PF   04-MAY-1994;   94WO-EP01409.
XX
PR   04-MAY-1994;   94WO-EP01409.
```

```
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ,  Waugh R;
XX
DR   WPI; 1995-404114/51.
XX
PT   Nucleic acid contg. plant U14 sequences and their derivs. - for
PT   regulating prodn. of ribosomal RNA, also as stabilisers for
PT   heterologous RNA
XX
PS   Example; Fig 7A-D; 57pp; English.
XX
CC   The present sequence shows the maize single U14 sequence U14.4,
CC   including flanking intergenic sequences and inverted repeats, as obtd.
CC   by sequencing of fragments of isolated genomic clones. Inverted repeats
CC   and Box C and D elements are shown above (see AAT08729-30). U14 genes
CC   are small nucleolar RNAs present in the genome as non-translated
CC   sequences. SnoRNAs are involved in processing of pre-ribosomal RNA
CC   transcripts and ribosome formation. U14snoRNA is required for a normal
CC   growth phenotype and for processing of pre-rRNA transcripts in yeast.
CC   Plant U14 sequences have been identified and characterised in the present
CC   invention, and may be intron encoded.
XX
SQ   Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;


  Query Match             60.0%;  Score 15;  DB 16;  Length 1683;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgct 15
           |||||||||||||||
Db     601 GTTGCTGGATCTGCT 587


RESULT   8
AAZ23024/c
ID   AAZ23024 standard; cDNA; 1689 BP.
XX
AC   AAZ23024;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Rat kd312 polypeptide encoding cDNA.
XX
KW   kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW   heart attack; head trauma; neurodegenerative disease; rat;
KW   Parkinson's disease; Alzheimer's disease; ss.
XX
OS   Rattus sp.
XX
PN   WO9950288-A2.
XX
PD   07-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06993.
XX
PR   31-MAR-1998;   98US-0053374.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Yen K;
XX
DR   WPI; 1999-601322/51.
DR   P-PSDB; AAY42694.
XX
PT   kd312 polypeptides useful for treating diseases and disorders
PT   associated with alterations in cell proliferation and cell death -
XX
PS   Claim 2; Fig 9; 95pp; English.
XX
CC   The invention provides nucleic acid molecules encoding human and rat
CC   kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC   recombinant methodology. The kd312 sequences, and the antibodies against
CC   the proteins may be used to treat or diagnose the presence or progression
CC   of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC   stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC   Parkinson's disease and Alzheimer's disease). The present sequence
CC   represents the rat kd312 cDNA sequence.
XX
SQ   Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 1689;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggatctgctgtttga 21
           |||||||||||||||
Db     495 GGATCTGCTGTTTGA 481


RESULT   9
AAC44132/c
ID   AAC44132 standard; DNA; 2022 BP.
XX
AC   AAC44132;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
```

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

```
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             60.0%;  Score 15;  DB 21;  Length 2022;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaagc 24
           |||||||||||||||
Db    1268 TCTGCTGTTTGAAGC 1254


RESULT  10
AAX04606/c
ID   AAX04606 standard; DNA; 2034 BP.
XX
AC   AAX04606;
XX
DT   07-APR-1999  (first entry)
XX
DE   Maize genomic clone MzU14.4 snoR cluster sequence.
XX
KW   Small nucleolar RNA gene; snoR; snoR1.4; maize; promoter;
KW   RNA production; RNA stability; small non-translated RNA;
KW   intergenic region; U49.1; SnoR2.4; U14.4; SnoR3.4; ds.
XX
OS   Zea sp.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..890
FT                   /*tag=  a
FT                   /note= "upstream region having promoter activity"
FT   misc_feature    891..983
FT                   /*tag=  b
FT                   /note= "SnoR1.4 coding sequence; no protein given in
FT                         specification"
FT   misc_feature    984..1100
FT                   /*tag=  c
FT                   /note= "intergenic region"
FT   misc_feature    1101..1294
FT                   /*tag=  d
FT                   /note= "U49.1 coding sequence; no protein given in
FT                         specification"
FT   misc_feature    1295..1361
FT                   /*tag=  e
FT                   /note= "intergenic region"
FT   misc_feature    1362..1512
FT                   /*tag=  f
FT                   /note= "SnoR2.4 coding sequence; no protein given in
FT                         specification"
FT   misc_feature    1513..1587
FT                   /*tag=  g
FT                   /note= "intergenic region"
FT   misc_feature    1588..1711
FT                   /*tag=  h
FT                   /note= "U14.4 coding sequence; no protein given in
FT                         specification"
FT   misc_feature    1712..1743
FT                   /*tag=  i
FT                   /note= "intergenic region"
FT   misc_feature    1744..1873
FT                   /*tag=  j
FT                   /note= "SnoR3.1 coding sequence; no protein given in
FT                         specification"
FT   misc_feature    1874..2034
FT                   /*tag=  k
FT                   /note= "3' flanking sequence"
XX
PN   EP887405-A1.
XX
PD   30-DEC-1998.
XX
PF   25-JUN-1997;   97EP-0401480.
XX
PR   25-JUN-1997;   97EP-0401480.
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ;
XX
DR   WPI; 1999-047872/05.
XX
PT   Producing stabilised RNA molecules  - using a DNA precursor
PT   comprising plant small nucleolar RNA stabilising sequences
XX
PS   Claim 46; Fig 18a-f; 98pp; English.
XX
CC   The present sequence represents the full sequence of the the small
CC   nucleolar RNA (snoRNA) cluster of maize genomic clone MzU14.4. The
CC   SnoR genes in the cluster are SnoR1.4, U49.1, SnoR2.4, U14.4, and
CC   SnoR3.1. Plant snoR genes enable production of multiple stabilised
CC   RNA from a single promoter in a splicing-independent manner,
CC   and allow the production of RNA from intron or non-intron sequences.
CC   The present sequence is used in the method of the invention. The
CC   specification describes a method for producing small non-translated
CC   RNA molecules, which comprises introducing a precursor DNA containing
CC   one or a cluster of coding sequences into a plant or yeast cell. Each
CC   coding sequence comprises a non-translated RNA sequence to be stabilised,
CC   which is capable of functionally interacting with a cellular component
CC   and stabilising sequences comprising at least one protein binding site.
CC   The stabilised RNA units are transcribed by the cell's endogenous
CC   processing machinery. The process is useful for qualitatively and
CC   quantitatively enhancing the production and stability of RNA molecules,
CC   especially non-translated RNA e.g ribozymes and antisense RNA. The
CC   nucleic acid molecules are useful as probes or primers for identification
CC   of snoRNA sequences and intergenic regions.
XX
SQ   Sequence 2034 BP; 449 A; 504 C; 435 G; 646 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 2034;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgct 15
           |||||||||||||||
Db     971 GTTGCTGGATCTGCT 957


RESULT  11
AAZ23025/c
ID   AAZ23025 standard; DNA; 3079 BP.
XX
AC   AAZ23025;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Rat kd312 genomic DNA sequence.
XX
KW   kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW   heart attack; head trauma; neurodegenerative disease; rat;
KW   Parkinson's disease; Alzheimer's disease; ss.
XX
OS   Rattus sp.
XX
PN   WO9950288-A2.
XX
PD   07-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06993.
XX
PR   31-MAR-1998;   98US-0053374.
XX
PA   (AMGE-) AMGEN INC.
XX
```

PI Yen K;
XX
DR WPI; 1999-601322/51.
DR P-PSDB; AAY42694.
XX
PT kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death -
XX
PS Claim 2; Fig 7; 85pp; English.
XX
CC The invention provides nucleic acid molecules encoding human and rat
CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC recombinant methodology. The kd312 sequences, and the antibodies against
CC the proteins may be used to treat or diagnose the presence or progression
CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC Parkinson's disease and Alzheimer's disease). The present sequence
CC represents the rat kd312 genomic DNA sequence.
XX
SQ Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ggatctgctgtttga 21
||||||||||||||||
Db 803 GGATCTGCTGTTTGA 789

RESULT 12
AAV49536
ID AAV49536 standard; DNA; 6703 BP.
XX
AC AAV49536;
XX
DT 16-NOV-1998 (first entry)
XX
DE Adenylate cyclase gene.
XX
KW ds; adenylate cyclase; recombinant microorganism; chemical production;
KW riboflavin; carotenoid.
XX
OS Ashbya gossypii.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..670
FT /*tag= a
FT CDS 671..6295
FT /*tag= b
FT /product= "Adenylate cyclase"
FT 3'UTR 6296..6703
FT /*tag= c
XX
PN WO9829538-A2.
XX
PD 09-JUL-1998.
XX
PF 29-DEC-1997; 97WO-EP07309.
XX
PR 31-DEC-1996; 97CH-0000016.
XX
PA (BADI) BASF AG.
XX
PI Althoefer H, Altmann-Joehl R, Philippsen P, Seulberger H;
XX
DR WPI; 1998-388120/33.
DR P-PSDB; AAW64518.
XX
PT New gene for adenylate cyclase from Ashbya gossypii - useful for
PT generating recombinant microorganisms with alteration in gene of
PT cAMP-dependent signalling pathway for increasing production of fine
PT chemicals
XX
PS Claim 1; Page 10-20; 33pp; German.
XX
CC The adenylate cyclase (AC) gene is used to generate recombinant
CC microorganisms for production of chemicals, particularly riboflavin but
CC also others useful in human or animal nutrition, cosmetics and medicines,
CC and also carotenoids for use as natural colourings. The modified
CC microorganisms produce chemicals at higher levels than wild-type
CC organisms. Particularly, for riboflavin, cells with a disrupted AC gene
CC are grown on glucose-containing medium.
XX
SQ Sequence 6703 BP; 1954 A; 1527 C; 1541 G; 1681 T; 0 other;

Query Match 60.0%; Score 15; DB 19; Length 6703;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctgt 17
|||||||||||||||
Db 6234 tgctggatctgctgt 6248

RESULT 13
AAZ12718
ID AAZ12718 standard; cDNA; 275 BP.
XX
AC AAZ12718;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:187.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 705; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell

```
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 275 BP; 65 A; 44 C; 48 G; 66 T; 52 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgct 15
           ||||||||||||||
Db     169 ttgctggatctgct 182


RESULT  14
AAF65487/c
ID   AAF65487 standard; cDNA; 336 BP.
XX
AC   AAF65487;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1243.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 722; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 336;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 atctgctgtttgaa 22
           ||||||||||||||
Db     159 ATCTGCTGTTTGAA 146


RESULT  15
AAV87319/c
ID   AAV87319 standard; cDNA; 379 BP.
XX
AC   AAV87319;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BP797.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 527; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
```

```
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 379 BP; 124 A; 68 C; 70 G; 116 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 379;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gatctgctgtttga 21
          ||||||||||||||
Db    129 GATCTGCTGTTTGA 116


RESULT  16
AAC26529/c
ID   AAC26529 standard; cDNA; 383 BP.
XX
AC   AAC26529;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30604.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30604; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 383 BP; 96 A; 103 C; 63 G; 119 T; 2 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgc 14
          ||||||||||||||
Db    263 GTTGCTGGATCTGC 250


RESULT  17
AAV01888/c
ID   AAV01888 standard; cDNA; 384 BP.
XX
AC   AAV01888;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human OTK27 gene.
XX
KW   Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW   brain specific nucleosome assembly protein; diagnosis; therapy;
KW   skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
KW   nel-related protein type 1; nel-related type 2; hereditary disease;
KW   cancer; OTK27; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..384
FT                   /*tag=  a
FT                   /product= "OTK27"
FT                   /note= "no stop codon given"
XX
PN   EP796913-A2.
XX
PD   24-SEP-1997.
XX
PF   19-MAR-1997;   97EP-0104842.
XX
PR   05-MAR-1997;   97JP-0069163.
PR   19-MAR-1996;   96JP-0063410.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Fujiwara T,  Horie M,  Watanabe T;
XX
DR   WPI; 1997-459830/43.
DR   P-PSDB; AAW37503.
XX
PT   Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT   useful for diagnosis or therapy of hereditary disease and cancer
XX
PS   Example 3; Page 35; 123pp; English.
XX
CC   The present sequence encodes a OTK27 isolated from a human foetal brain
CC   cDNA library. The nucleotide or amino acid sequences are useful for
CC   in-vitro diagnosis of hereditary diseases and cancer and for preparation
CC   of pharmaceuticals.
XX
SQ   Sequence 384 BP; 94 A; 111 C; 98 G; 81 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 81;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 tggatctgctgttt 19
          ||||||||||||||
Db    350 TGGATCTGCTGTTT 337


RESULT  18
AAC99862
ID   AAC99862 standard; cDNA; 612 BP.
XX
AC   AAC99862;
XX
```

DT 13-MAR-2001 (first entry)
XX
DE Human secreted protein gene 45 SEQ ID NO:55.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200070042-A1.
XX
PD 23-NOV-2000.
XX
PF 11-MAY-2000; 2000WO-US12788.
XX
PR 13-MAY-1999; 99US-0134068.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
XX
DR WPI; 2000-679828/66.
DR P-PSDB; AAB56121.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 881; 1065pp; English.
XX
CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and opthalmalogical. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention.
XX
SQ Sequence 612 BP; 199 A; 139 C; 139 G; 135 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 612;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgctggatctgct 15
||||||||||||||
Db 372 ttgctggatctgct 385

RESULT 19
AAA44124/c
ID AAA44124 standard; cDNA; 619 BP.
XX
AC AAA44124;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:699.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24206.
XX
PR 15-OCT-1998; 98US-0104436.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 383; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX

SQ Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 619;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atctgctgtttgaa 22
||||||||||||||
Db 43 ATCTGCTGTTTGAA 30

RESULT 20
AAF14018/c
ID AAF14018 standard; cDNA; 707 BP.
XX
AC AAF14018;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:6541.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2678; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 707 BP; 154 A; 241 C; 156 G; 156 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 707;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ctggatctgctgtt 18
||||||||||||||
Db 693 CTGGATCTGCTGTT 680

RESULT 21
AAQ92502
ID AAQ92502 standard; cDNA; 735 BP.
XX
AC AAQ92502;
XX
DT 07-FEB-1996 (first entry)
XX
DE Mouse antibody F4-7 heavy chain variable region coding sequence.
XX
KW Primer; amplification; PCR; mouse; kappa chain; heavy chain; Fab;
KW antibody; immunotolerance; animal; variegated display library;
KW variable region; antigen; immunorecessive; cell surface marker; foetal;
KW cancer; stem cell; variant; therapy; Alzheimer's disease; hybridoma;
KW familial hypercholesterolaemia; binding affinity; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 67..735
FT /*tag= a
FT /product= antibody F4-7 heavy chain variable region
XX
PN WO9515982-A2.
XX
PD 15-JUN-1995.
XX
PF 08-DEC-1994; 94WO-US14106.
XX
PR 06-DEC-1994; 94US-0350400.
PR 08-DEC-1993; 93US-0164022.
XX
PA (GENZ) GENZYME CORP.
XX
PI Barsomian G, Copeland DP, Hillhouse D, Johnson T;
XX
DR WPI; 1995-224291/29.
DR P-PSDB; AAR75458.
XX
PT Generating new antibodies specific for immunorecessive epitopes -
PT by selection from variegated V gene library cloned from
PT immuno:tolerance derived antibody repertoire, useful in diagnosis,
PT purifcn. and therapy, e.g. of cancer
XX
PS Disclosure; Page 77-78; 109pp; English.
XX
CC The coding sequence of the heavy chain variable region from the mouse
CC antibody F4-7. This sequence was isolated from a variegated display
CC library (VDL) of variable regions derived from a repertoire of
CC antibodies from an immunotolerised animal. The VDL is generated by PCR
CC amplifying the variable regions from the antibody coding sequences using
CC the primers AAQ74153-74. The variable regions, esp the complementarity
CC determining regions (CDR; see AAR75462-93 for examples of CDRs) from the
CC immunotolerant animals' antibodies are used to construct an antibody
CC against a immunorecessive antigen e.g. a cell surface marker on a foetal,
CC cancer or stem cell, which can differentiate between variant or related
CC forms of the antigen. The antibodies generated can be used in the
CC diagnosis, e.g. detection of the immunorecessive antigen, or in therapy

CC e.g. of cancer, Alzheimer's disease or familial hypercholesterolaemia.
CC The method of production of the antibody allows rapid and sensitive
CC isolation of antibodies that would be difficult to isolate by standard
CC methods. The antibodies produced have greater binding affinity than
CC those produced by combinatorial/hybridoma methods.
XX
SQ Sequence 735 BP; 172 A; 213 C; 188 G; 162 T; 0 other;

Query Match 56.0%; Score 14; DB 16; Length 735;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctg 16
||||||||||||||
Db 465 tgctggatctgctg 478

RESULT 22
AAZ00347/c
ID AAZ00347 standard; DNA; 753 BP.
XX
AC AAZ00347;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleotide sequence of human hsFATP3.
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO9936537-A2.
XX
PD 22-JUL-1999.
XX
PF 14-JAN-1999; 99WO-US00182.
XX
PR 14-JAN-1999; 99US-0232201.
PR 15-JAN-1998; 98US-0071374.
PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14937.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Example 1; Fig 16; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety.,DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 436 TCTGCTGTTTGAAG 423

RESULT 23
AAF15012/c
ID AAF15012 standard; cDNA; 769 BP.
XX
AC AAF15012;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7535.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3043-3044; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an

```
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 769 BP; 203 A; 234 C; 203 G; 114 T; 15 other;


  Query Match            56.0%;  Score 14;  DB 21;  Length 769;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db    284 TGCTGGATCTGCTG 271


RESULT  24
AAZ06233
ID   AAZ06233 standard; DNA; 891 BP.
XX
AC   AAZ06233;
XX
DT   30-SEP-1999  (first entry)
XX
DE   Human secreted protein gene No. 15.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9935158-A1.
XX
PD   15-JUL-1999.
XX
PF   06-JAN-1999;   99WO-US00108.
XX
PR   07-JAN-1998;   98US-0070704.
PR   07-JAN-1998;   98US-0070657.
PR   07-JAN-1998;   98US-0070658.
PR   07-JAN-1998;   98US-0070692.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Duan RD,  Ebner R,  Lafleur DW,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR;
XX
DR   WPI; 1999-444190/37.
DR   P-PSDB; AAY38400, AAY38453.
XX
PT   New isolated human genes and the secreted polypeptides they encode
XX
PS   Claim 1; Page 158; 227pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a
CC   secreted human protein. The gene number is given in the descriptor line.
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. AAZ06210) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 36 novel genes and their fragments (nucleic
CC   acid sequences: AAZ06219-Z06263; amino acid sequences AAY38386-Y38498)
CC   which are useful for preventing, treating or ameliorating  medical
CC   conditions e.g. by protein or gene therapy.  Also, pathological
CC   conditions can be diagnosed by determining the amount of the new
CC   polypeptides in a sample or by determining the presence of mutations in
CC   the new polynucleotides. Specific uses are described for each of the 36
CC   polynucleotides, based on which tissues they are most highly expressed in
CC   (see AAZ06219 for described uses).
XX
SQ   Sequence 891 BP; 207 A; 214 C; 220 G; 250 T; 0 other;


  Query Match            56.0%;  Score 14;  DB 20;  Length 891;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgct 15
          ||||||||||||||
Db    727 ttgctggatctgct 740


RESULT  25
AAX82834/c
ID   AAX82834 standard; DNA; 1044 BP.
XX
AC   AAX82834;
XX
DT   30-JUN-2000  (first entry)
XX
DE   L-asparaginase DNA.
XX
KW   L-asparaginase; fermentation; ss.
XX
OS   Unidentified.
XX
PN   CN1237633-A.
XX
PD   08-DEC-1999.
XX
PF   01-JUN-1998;   98CN-0102046.
XX
PR   01-JUN-1998;   98CN-0102046.
XX
PA   (MICR-) INST MICROBIOLOGY CHINESE ACAD SCI.
XX
PI   Qian S,  Wang Y,  Meng G;
XX
DR   WPI; 2000-351194/31.
DR   P-PSDB; AAW90867.
XX
PT   New structured L-asparaginase bacterial host cell and its production
PT   culture -
XX
PS   Disclosure; Fig 2; 12pp; Chinese.
XX
CC   This invention describes a novel L-asparaginase-producing recombinant
CC   bacteria and its fermentation culture method. The new cell is produced
CC   using standard recombinant techniques. The enzyme gene expressed in
CC   the cell is different from the reported enzyme gene in sequence, and
CC   its individual amino acid composition is also different. The culture
CC   medium adopted for culturing the new bacterial cells is cheap and easily
CC   available, and the fermentation process used to grow the cells and
CC   express the enzyme is simple and easy to implement. This sequence
CC   encodes an L-asparaginase which is described in the method of the
CC   invention.
XX
SQ   Sequence 1044 BP; 274 A; 281 C; 272 G; 217 T; 0 other;


  Query Match            56.0%;  Score 14;  DB 21;  Length 1044;
  Best Local Similarity  100.0%;  Pred. No. 82;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   1028 TGCTGGATCTGCTG 1015
```

```
RESULT  26
AAZ09784
ID   AAZ09784 standard; DNA; 1155 BP.
XX
AC   AAZ09784;
XX
DT   23-NOV-1999  (first entry)
XX
DE   E. coli S-adenosyl methionine synthase SAM DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1155
FT                   /*tag=  a
FT                   /product= "S-adenosyl methionine synthase"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33263.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Claim 1; Page 8-10; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the Escherichia coli SAM protein which is used in the method of the
CC   invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   1031 tgctggatctgctg 1044


RESULT  27
AAA99433
ID   AAA99433 standard; DNA; 1155 BP.
XX
AC   AAA99433;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
XX
KW   L-methionine production; Met; Escherichia coli; ds.
XX
OS   Escherichia coli.
XX
PN   JP2000139471-A.
XX
PD   23-MAY-2000.
XX
PF   17-NOV-1998;   98JP-0326717.
XX
PR   17-NOV-1998;   98JP-0326717.
XX
PA   (AJIN ) AJINOMOTO KK.
XX
DR   WPI; 2001-018703/03.
DR   P-PSDB; AAB26814.
XX
PT   Fermentative process for preparation of L-methionine (Met), comprises
PT   using a modified Met producing microorganism, particularly an
PT   Escherichia species -
XX
PS   Example 2; Page 17-18; 23pp; Japanese.
XX
CC   The invention relates to a process for preparing L-methionine (Met)
CC   using a modified Met producing organism. The modified microorganism has
CC   its Met biosynthetic system repressor deleted, and has enhanced
CC   homoserine transsuccunylase activity. The organism optionally has an
CC   attenuated intracellular S-adenosyl methionine synthetase (SAM). The
CC   process is used for the preparation of L-methionine. Sequences AAA99416
CC   and AAA99433 encode proteins represented in AAB26816 and AAB26814
CC   respectively, which are involved in the Escherichia coli Met biosynthetic
CC   pathway. AAA99417-A99441 (excluding AAA99433) represent primers, all
CC   sequences are used in the production of the modified Met producing
CC   organism of the invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 82;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   1031 tgctggatctgctg 1044


RESULT  28
AAZ00358/c
ID   AAZ00358 standard; DNA; 1173 BP.
XX
AC   AAZ00358;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of human hsFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
```

PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14948.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Claim 31; Fig 48; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db     853 TCTGCTGTTTGAAG 840
```

RESULT 29
AAC44861
ID AAC44861 standard; DNA; 1220 BP.
XX
AC AAC44861;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44409.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.0%; Score 14; DB 21; Length 1220;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     317 tgctggatctgctg 330
```

RESULT 30
AAA81673/c
ID AAA81673 standard; DNA; 1367 BP.
XX
AC AAA81673;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_220 SEQ ID NO:220.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX

```
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 1535; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 1367 BP; 412 A; 313 C; 290 G; 352 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1367;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 gctggatctgctgt 17
          ||||||||||||||
Db   1013 GCTGGATCTGCTGT 1000


RESULT  31
AAV01889/c
ID   AAV01889 standard; DNA; 1493 BP.
XX
AC   AAV01889;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human OTK27 gene.
XX
KW   Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW   brain specific nucleosome assembly protein; diagnosis; therapy;
KW   skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
KW   nel-related protein type 1; nel-related type 2; hereditary disease;
KW   cancer; OTK27; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             95..481
FT                   /*tag=  a
FT                   /product= "OTK27"
XX
PN   EP796913-A2.
XX
PD   24-SEP-1997.
XX
PF   19-MAR-1997;   97EP-0104842.
XX
PR   05-MAR-1997;   97JP-0069163.
PR   19-MAR-1996;   96JP-0063410.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Fujiwara T,  Horie M,  Watanabe T;
XX
DR   WPI; 1997-459830/43.
DR   P-PSDB; AAW37503.
XX
PT   Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT   useful for diagnosis or therapy of hereditary disease and cancer
XX
PS   Example 3; Page 36-37; 123pp; English.
XX
CC   The present sequence encodes a OTK27 isolated from a human foetal brain
CC   cDNA library. The nucleotide or amino acid sequences are useful for
CC   in-vitro diagnosis of hereditary diseases and cancer and for preparation
CC   of pharmaceuticals.
XX
SQ   Sequence 1493 BP; 339 A; 363 C; 394 G; 397 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 tggatctgctgttt 19
          ||||||||||||||
Db    444 TGGATCTGCTGTTT 431


RESULT  32
AAD03029/c
ID   AAD03029 standard; DNA; 1676 BP.
XX
AC   AAD03029;
XX
DT   31-MAY-2001  (first entry)
XX
DE   Flax 2S storage protein DNA.
XX
KW   Flax; seed-specific promoter; storage protein; seed oil; vaccine;
KW   protein expression; anticoagulant; cytokine; growth factor; pectinase;
KW   interleukin; alpha-1-antitrypsin; anti-obesity protein; haemoglobin;
KW   serum albumin; insulin; lactoferrin; myoglobin; pulmonary surfactant;
KW   alpha amylase; ds.
XX
OS   Linum usitatissimum.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (1..18)
FT                   /*tag=  a
FT                   /bound_moiety= "5' PCR primer(1)"
FT   repeat_unit     25..36
FT                   /*tag=  b
FT                   /label= AT_rich_repeat
FT   repeat_unit     97..108
FT                   /*tag=  c
FT                   /label= AT_rich_repeat
FT   repeat_unit     167..180
FT                   /*tag=  d
FT                   /label= AT_rich_repeat
FT   repeat_unit     241..247
FT                   /*tag=  e
FT                   /label= RY_like_repeat
FT   misc_signal     275..279
FT                   /*tag=  f
FT                   /label= G-box
FT   misc_feature    285..290
FT                   /*tag=  g
```

```
FT                   /note= "Seed specific box-like motif"
FT   TATA_signal     327..333
FT                   /*tag=  h
FT   primer_bind     381..398
FT                   /*tag=  i
FT                   /bound_moiety= "3' PCR primer(1)"
FT   CDS             418..942
FT                   /*tag=  j
FT                   /product= "Flax 2S storage protein"
FT   primer_bind     complement (943..957)
FT                   /*tag=  k
FT                   /bound_moiety= "5' PCR primer(2)"
FT   primer_bind     1158..1676
FT                   /*tag=  l
FT                   /bound_moiety= "3' PCR primer(2)"
XX
PN   WO200116340-A1.
XX
PD   08-MAR-2001.
XX
PF   25-AUG-2000; 2000WO-CA00988.
XX
PR   27-AUG-1999;   99US-0151044.
PR   27-OCT-1999;   99US-0161722.
PR   30-MAY-2000; 2000CA-2310304.
XX
PA   (SEMB-) SEMBIOSYS GENETICS INC.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Chaudhary S,  Van Rooijen G,  Moloney MM,  Singh S;
XX
DR   WPI; 2001-226693/23.
DR   P-PSDB; AAY72901.
XX
PT   Expressing non-native genes in flax seeds and seeds of other plant
PT   species for altering the seed oil and protein composition in the seeds,
PT   comprises using seed-specific promoters obtained from flax  -
XX
PS   Claim 14; Fig 3; 68pp; English.
XX
CC   The present sequence is a DNA encoding flax 2S storage protein. The
CC   present invention relates to a method for expressing non-native genes in
CC   flax seeds. The method comprises introducing a chimeric nucleic acid
CC   construct containing a seed-specific promoter obtained from flax and a
CC   nucleic acid which is non-native to the promoter, into a flax plant cell
CC   and growing the plant cell into a mature flax plant capable of setting
CC   seed, where the nucleic acid is expressed in the seed under the control
CC   of the promoter. The seed-specific promoters obtained from flax are
CC   useful for modifying the protein, oil or polysaccharide composition of
CC   the flax seeds and seeds of other plant species. The promoters facilitate
CC   expression of proteins, including sulphur-rich protein that are found in
CC   lupins or Brazil nuts in a seed deficient in sulphurous amino acids,
CC   peptides having pharmaceutical value such as anticoagulants, antibodies,
CC   vaccines, cytokines, growth factors, interleukins, mammalian proteins,
CC   including alpha-1-antitrypsin, anti-obesity proteins, haemoglobin, blood
CC   proteins, human serum albumin, insulin, lactoferrin, myoglobin, pulmonary
CC   surfactants and proteins of industrial value such as alpha-amylase,
CC   arabinase, amyloglucosidase, catalase, cellobiohydrolase, pectinases,
CC   phytase, papain and xylanase.
XX
SQ   Sequence 1676 BP; 516 A; 367 C; 442 G; 351 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1676;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     584 TGCTGGATCTGCTG 571


RESULT  33
AAT96346/c
ID   AAT96346 standard; cDNA; 1848 BP.
XX
AC   AAT96346;
XX
DT   08-APR-1998  (first entry)
XX
DE   Chimeric gene containing anti-asparaginase MAb light and heavy chain.
XX
KW   Immunoglobulin; Ig; heavy chain; variable region; murine; human;
KW   asparaginase II; monoclonal antibody; MAb; light chain;
KW   recombinant chimeric polypeptide; ss.
XX
OS   Synthetic.
OS   Chimeric - Homo sapiens.
OS   Chimeric - Mus sp.
XX
PN   US5686579-A.
XX
PD   11-NOV-1997.
XX
PF   23-MAY-1995;   95US-0447422.
XX
PR   22-JUN-1993;   93US-0081410.
PR   21-JUN-1988;   88US-0205748.
PR   31-AUG-1992;   92US-0938505.
PR   23-MAY-1995;   95US-0447422.
XX
PA   (HYBR-) HYBRISENS LTD.
XX
PI   Ramjeesingh M,  Rothstein A,  Shami EY;
XX
DR   WPI; 1997-558200/51.
XX
PT   Self-protecting chimeric polypeptide comprising biologically active
PT   sequence and single-chain antibody sequence - has resistance to e.g.
PT   disrupting temperature, presence of proteolytic enzymes, etc.
XX
PS   Example 2; Columns 27-30; 29pp; English.
XX
CC   The present sequence is a chimeric gene containing the cDNA for
CC   the immunoglobulin (Ig) heavy and light chain variable regions of a
CC   murine anti-asparaginase II monoclonal antibody (MAb), and human
CC   asparaginase II. The gene was used in the preparation of a novel
CC   recombinant chimeric polypeptide, comprising a 1st region
CC   comprising a biologically active domain and another domain
CC   containing an epitope, linked via a polypeptide to a 2nd region
CC   including a single chain antibody (SCA) having the light and heavy
CC   chains of an antibody variable region which specifically binds the
CC   epitope in the 1st region. The chimeric polypeptide assumes a
CC   conformation in which the SCA is bound to the epitope of the 1st
CC   region and protects its biological activity from deactivation by
CC   denaturing temperatures or pH conditions, proteolytic enzymes,
CC   oxidising agents or alcohol. The regions of the chimeric
CC   polypeptide interact to form a structure analogous to an
CC   antibody-antigen complex. A L-asparaginase-SCA fusion protein of
CC   the above type has better trypsin resistance than free
CC   L-asparaginase.
XX
SQ   Sequence 1848 BP; 435 A; 498 C; 517 G; 398 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 1848;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    1829 TGCTGGATCTGCTG 1816


RESULT  34
AAZ00367/c
```

```
ID   AAZ00367 standard; DNA; 1998 BP.
XX
AC   AAZ00367;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of murine mmFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Mus sp.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14957.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Example 1; Fig 66; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC   uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC   muscle or liver by administration of a complex of the agent and a FATP6
CC   binding moiety. DNA encoding FATP proteins can be used as a reference
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
SQ   Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 83;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1690 TCTGCTGTTTGAAG 1677


RESULT  35
AAZ30162
ID   AAZ30162 standard; cDNA; 2080 BP.
XX
AC   AAZ30162;
XX
DT   26-JAN-2000  (first entry)
XX
DE   cDNA encoding a portion of the starch R1 phosphorylation protein.
XX
KW   Starch R1 phosphorylation protein; starch degradation;
KW   plant starch biosynthesis; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1748
FT                   /*tag=  a
XX
PN   WO9953072-A1.
XX
PD   21-OCT-1999.
XX
PF   08-APR-1999;   99WO-US07639.
XX
PR   09-APR-1998;   98US-0081143.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cressman RF,  Allen SM;
XX
DR   WPI; 1999-620435/53.
DR   P-PSDB; AAY43629.
XX
PT   New starch R1 phosphorylation protein homologues  -
XX
PS   Claim 2; Page 41-42; 54pp; English.
XX
CC   The present sequence encodes a starch R1 phosphorylation protein.
CC   The protein is involved in starch degradation. The specification
CC   also describes a chimeric gene encoding all or a portion of the
CC   starch R1 phosphorylation protein, in the sense or antisense
CC   orientation, where expression of the chimeric gene results in
CC   production of altered levels of starch R1 phosphorylation protein
CC   in transformed plants or cells. The protein facilitates studies of
CC   starch degradation in plants, and provides tools for the genetic
CC   manipulation of starch biosynthesis in plants.
XX
SQ   Sequence 2080 BP; 623 A; 385 C; 467 G; 605 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 2080;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     290 tgctggatctgctg 303


RESULT  36
AAZ00343/c
ID   AAZ00343 standard; DNA; 2087 BP.
XX
AC   AAZ00343;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of murine mmFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Mus musculus.
XX
PN   WO9936537-A2.
```

```
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14933.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Example 1; Fig 8; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC   uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC   muscle or liver by administration of a complex of the agent and a FATP6
CC   binding moiety. DNA encoding FATP proteins can be used as a reference
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
SQ   Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  37
AAZ00379/c
ID   AAZ00379 standard; DNA; 2166 BP.
XX
AC   AAZ00379;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of human hsFATP5 gene.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14969.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Claim 31; Fig 94A-B; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC   uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC   muscle or liver by administration of a complex of the agent and a FATP6
CC   binding moiety. DNA encoding FATP proteins can be used as a reference
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
SQ   Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 2166;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1987 TCTGCTGTTTGAAG 1974


RESULT  38
AAF27417/c
ID   AAF27417 standard; cDNA; 2405 BP.
XX
AC   AAF27417;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human fatty acid transporter PSEC67 cDNA.
XX
KW   Human; fatty acid transporter; PSEC67; long-chain fatty acid uptake;
KW   oleic acid; drug screening; gene therapy; metabolic disorder;
KW   cardiomyopathy; skeletal muscle disorders; renal failure; ss.
XX
OS   Homo sapiens.
XX
PN   WO200104301-A1.
XX
```

PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-JP04549.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 18-OCT-1999; 99US-0159586.
PR 25-APR-2000; 2000JP-0128993.
XX
PA (HELI-) HELIX RES INST.
XX
PI Morikawa N, Masuho Y, Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
DR WPI; 2001-138349/14.
DR P-PSDB; AAB60388.
XX
PT Fatty acid transporter protein and encoded gene PSEC67 cloned from
PT human cDNA library, with activity of oleic acid incorporation, useful
PT as target molecule of preventives or remedies of fatty-acid metabolic
PT disorders -
XX
PS Claim 1; Page 44-48; 58pp; Japanese.
XX
CC The invention relates to a novel human fatty acid transporter,
CC PSEC67 (AAB60388), and to cDNA encoding it (AAF27417). PSEC67 is
CC responsible for the uptake of oleic acid into cells. The invention
CC also relates to vectors and host cells comprising a PSEC67
CC nucleic acid; the recombinant production of PSEC67; an antibody
CC against PSEC67; methods of screening for compounds which can regulate
CC the uptake of long-chain fatty acids into cells; and the compounds thus
CC identified. The PSEC67 protein and the gene encoding it are useful as
CC targets for the treatment or prevention of diseases associated with
CC defective metabolism of long-chain fatty acids. Such diseases include
CC cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
CC nucleic acids may also be used in gene therapy for such disorders.
CC The present sequence represents cDNA encoding PSEC67.
XX
SQ Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2405;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 2111 TCTGCTGTTTGAAG 2098

RESULT 39
AAZ33977/c
ID AAZ33977 standard; cDNA; 2574 BP.
XX
AC AAZ33977;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO703 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.

PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41699.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 2; Fig 38; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2574;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

RESULT 40
AAC78481/c
ID AAC78481 standard; cDNA; 2574 BP.
XX
AC AAC78481;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO703 (UNQ367) nucleotide sequence SEQ ID NO:101.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
DR WPI; 2000-611443/58.
DR P-PSDB; AAB44255.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
PS Claim 2; Fig 38; 636pp; English.
XX
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

RESULT 41
AAC58239/c
ID AAC58239 standard; cDNA; 2574 BP.
XX
AC AAC58239;
XX
DT 25-JAN-2001 (first entry)
XX
DE Human PRO703 nucleotide sequence SEQ ID NO:28.
XX
KW Human; tumour; diagnosis; neoplastic disease; identification; cancer;
KW tumourigenesis; detection; neoplastic cell growth; proliferation;
KW cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
KW immunological disorder; ss.
XX

OS Homo sapiens.
XX
PN WO200053754-A1.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000WO-US00277.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 05-OCT-1999; 99WO-US23089.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;
PI Wood WI;
XX
DR WPI; 2000-572269/53.
DR P-PSDB; AAB24054.
XX
PT New isolated antibody for use in compositions and methods for the
PT diagnosis and treatment of neoplastic cell growth and proliferation in
PT mammals, including humans, and in monitoring tumor treatment -
XX
PS Claim 50; Fig 28; 195pp; English.
XX
CC The present invention describes an isolated antibody (Ab) that binds to
CC one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
CC PRO237, PRO324, PRO351, PRO362, PRO615, PRO531, PRO538, PRO3664, PRO618,
CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
CC and methods for the diagnosis and treatment of neoplastic cell growth
CC and proliferation in mammals, including humans. Genes and polypeptides
CC encoded by them, that are amplified in the genome of a tumour cell, can
CC be identified and are useful targets for the treatment and prevention of
CC certain cancers and may be used to monitor tumour treatment. Compounds
CC that inhibit the expression or activity of the identified polypeptides
CC can be identified and used as antagonists. Benign or malignant tumours,
CC inflammatory disorders and immunological disorders can be treated.
CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

RESULT 42
AAZ25020/c
ID AAZ25020 standard; cDNA; 2669 BP.
XX
AC AAZ25020;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human H-beta 58 family CBLALC05 nucleotide sequence.
XX
KW Human; H-beta 58 family; CBLALC05; cancer; AIDS; metabolic disorder;
KW neurological disease; developmental abnormality; diagnosis;
KW spontaneous abortion; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 80..1063
FT /*tag= a
FT /product= "CBLALC05"
FT /note= "an H-beta 58 family protein"
XX
PN WO9947663-A1.
XX
PD 23-SEP-1999.
XX
PF 18-MAR-1998; 98WO-CN00042.
XX
PR 18-MAR-1998; 98WO-CN00042.
XX
PA (UYSH-) UNIV SHANGHAI SECOND MEDICAL.
XX
PI Zhang Q, Kan L, He K, Shen Y;
XX
DR WPI; 1999-571836/48.
DR P-PSDB; AAY41679.
XX
PT CBLALC05, a homologue of murine H-beta 58, and related polynucleotides
XX
PS Claim 4; Page 9-11; 33pp; English.
XX
CC The present sequence encodes a human H-beta 58 family protein designated
CC CBLALC05. CBLALC05 is homologous to the murine H-beta 58 protein. The
CC CBLALC05 polynucleotide (I), vectors containing (I) and recombinant host
CC cells are useful for recombinant production of CBLALC05. Both (I),
CC CBLALC05 and antibodies against CBLALC05 are useful as research reagents,
CC for screening assays and in diagnostic assays. Antagonists and agonists
CC of CBLALC05 can be used to inhibit or enhance, respectively, the activity
CC of CBLALC05 or expression of (I). Anti-CBLALC05 antibodies and CBLALC05
CC or its fragments can be used in vaccines. In particular, the proteins,
CC antibodies, agonists and antagonists can be used for treating, e.g.
CC cancer, AIDS, metabolic disorders, neurological disease, developmental
CC abnormalities and spontaneous abortions, related to both an excess and
CC insufficient amounts of CBLALC05.
XX
SQ Sequence 2669 BP; 864 A; 419 C; 530 G; 856 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2669;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ctgctgtttgaagc 24
||||||||||||||
Db 949 CTGCTGTTTGAAGC 936

RESULT 43
AAC98021/c
ID AAC98021 standard; cDNA; 2675 BP.
XX
AC AAC98021;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:31.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;

```
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53264.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 490-491; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 2675 BP; 856 A; 425 C; 532 G; 855 T; 7 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 2675;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 ctgctgtttgaagc 24
           ||||||||||||||
Db     972 CTGCTGTTTGAAGC 959


RESULT  44
AAQ05686/c
ID   AAQ05686 standard; DNA; 3243 BP.
XX
AC   AAQ05686;
XX
DT   03-JAN-1991  (first entry)
XX
DE   Islets of Langerhans cell clone ICA12.3 (ATCC 40703).
XX
KW   Pancreatic Islet Cell Antigen (ICA) 12.3; ATCC 40703;
KW   insulin-dependent (Type I) diabetes mellitus; ss;
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             692..1618
FT                   /*tag= 'a
XX
PN   EP383129-A.
XX
PD   22-AUG-1990.
XX
PF   03-FEB-1990;   90EP-0102179.
XX
PR   04-DEC-1989;   89US-0441703.
PR   17-FEB-1989;   89US-0312543.
XX
PA   (MOLE-) MOLECULAR DIAGNOSTI.
XX
PI   Rabin D;
XX
DR   WPI; 1990-255450/34.
DR   P-PSDB; AAR06467.
XX
PT   Pancreatic islet cell antigens obtd. by mol. cloning - used in
PT   research, diagnosis and treatment of insulin-dependent (Type I)
PT   diabetes mellitus.
XX
PS   Disclosure; Fig 13; 45pp; English.
XX
CC   A cDNA library was constructed from RNA extracted from purified
CC   human islets of Langerhans cells. The library was screened with
CC   sera from autoimmune patients with Type I diabetes mellitus.
CC   ICA 12.3 is one of nine clones found to react predominantly with
CC   diabetic sera. The protein expressed by the clone was analysed by
CC   expressing the clones in E.coli hosts. The protein or peptide
CC   fragments of it can be used as immunoassay reagents in the
CC   presymptomatic diagnosis of insulin-dependent diabetes mellitus
CC   (IDDM).
CC   They may also be used to identify or mark T- or B-cells involved in
CC   IDDM or to bind or block these same cells for therapy.
CC   See also AAQ05681-Q05685 and AAQ05687-Q05688.
XX
SQ   Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 11;  Length 3243;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     413 TGCTGGATCTGCTG 400


RESULT  45
AAZ32338/c
ID   AAZ32338 standard; cDNA; 3243 BP.
XX
AC   AAZ32338;
XX
DT   21-JAN-2000  (first entry)
XX
DE   Human pancreatic islet cell antigen ICA-12.3 ATCC 40703 cDNA sequence.
XX
KW   Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
KW   insulin dependent type I diabetes mellitus; diagnosis; detection;
KW   immunoglobulin; T-cell; B-cell; antibody binding; ss.
XX
OS   Homo sapiens.
XX
PN   US5981700-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0468579.
XX
PR   05-MAY-1994;   94US-0239276.
PR   14-JUN-1991;   91US-0715181.
PR   08-JUN-1992;   92US-0872646.
PR   17-FEB-1989;   89US-0312543.
```

PR 04-DEC-1989; 89US-0441703.
XX
PA (FARB) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX
PI Rabin DU;
XX
DR WPI; 1999-633370/54.
DR P-PSDB; AAY49852.
XX
PT New pancreatic islet cell antigens, useful in the diagnosis of
PT insulin-dependent (Type 1) diabetes mellitus -
XX
PS Claim 1; Column 35-42; 61pp; English.
XX
CC The present sequence encodes a human pancreatic islet cell antigen (ICA)
CC that binds with antibodies found in the sera of patients afflicted with
CC insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
CC present invention are encoded by the DNA insert of a recombinant cloning
CC vehicle selected from ATCC 40550 (AAZ32333, encoding AAY49847), ATCC
CC 40553 (AAZ32334, encoding AAY49848), ATCC 40554 (AAZ32335, encoding
CC AAY49849), ATCC 40551 (AAZ32336, encoding AAY49850), ATCC 40552
CC (AAZ32337, encoding AAY49851), ATCC 40703 (AAZ32338, encoding AAY49852),
CC ATCC 40704 (AAZ32339, encoding AAY49853), ATCC 40705 (AAZ32340), ATCC
CC 40706 (AAZ32341, encoding AAY49854) and ATCC 75030 (AAZ32342, encoding
CC AAY49855). ICA proteins and their peptide fragments can be used in the
CC diagnosis of IDDM and in detecting or blocking human immunoglobulin,
CC T-cells or B-cells involved in IDDM.
XX
SQ Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 3243;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctg 16
||||||||||||||
Db 413 TGCTGGATCTGCTG 400

Search completed: October 9, 2001, 15:52:08
Job time: 15089 sec